UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

PROBLEMAS INVERSOS Y CÁLCULO DE VARIACIONES: APLICACIONES A LA BIOLOGÍA Y AL ANÁLISIS DE IMÁGENES

TESIS PARA OPTAR AL TÍTULO DE GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MODELACIÓN MATEMÁTICA

RODRIGO ANTONIO LECAROS LIRA

PROFESORES GUÍAS:
CARLOS CONCA ROSENDE
JAIME H. ORTEGA PALMA

PROFESOR CO-GUÍA:
LIONEL ROSIER

MIEMBROS DE LA COMISIÓN:
MATÍAS COURDURIER BETTANCOURT
ZIOMARA GERDTZEN HAKIM

Este trabajo ha sido parcialmente financiado por Conicyt

SANTIAGO DE CHILE
AGOSTO 2012

Resumen

El tema central de esta tesis, para optar al grado de Doctor en Ciencias de la Ingeniería con mención en Modelamiento Matemático, es el estudio de los llamados problemas inversos, el modelamiento matemático de fenómenos de interacción fluido estructura y la implementación numérica de un algoritmo de segmentación.

Este trabajo se divide en tres partes. En primer lugar (Capítulo 2), estudiamos un problema inverso en biologia. Nos interesamos en la posibilidad de recuperar la densidad de un canal iónico (canal **CNG**) en el cilio olfatorio, sólo midiendo la corriente producida al activar el sistema mediante la difusión de un agente en el interior del cilio. Este problema es modelado como un problema inverso en que interviene una ecuación de Fredholm lineal, de primer tipo, con un núcleo difusivo. Consideramos una aproximación para el núcleo del operador, obteniendo los siguientes resultados:

a) Problema de identificabilidad: Bajo diferentes regularidades de la función de densidad, logramos establecer la inyectividad del operador bajo estudio.

b) Problema de estabilidad: Tanto en el caso general, usando la transformada de Mellin, como en un caso particular de los parámetros del problema, logramos establecer la estabilidad, bajo diferentes normas, es decir, la continuidad del inverso del operador.

d) Problema de reconstrucción: En el caso de una elección apropiada de los parámetros del problema, logramos construir la función inversa.

e) Reconstrucción numerica: Utilizando la inversa, construida en la parte anterior, implementamos un algoritmo numerico de reconstrucción de la densidad.

Por otra parte, en el Capítulo 3 investigamos la controlabilidad de un submarino inmerso en un volumen infinito de un fluido potencial. Consideramos como control, el flujo del fluido a través de una parte de la frontera del móvil. Se llega a un sistema finito dimensional, similar al sistema de Kirchhoff, en el cual los controles aparecen a través de un término lineal (con derivada temporal) y uno bilineal. Aplicando el método del retorno, establecimos bajo ciertas condiciones geométricas un resultado de controlabilidad local para la posición y velocidad del vehículo sólo con cuatro controles.

Finalmente, en el Capítulo 4 implementamos un algoritmo numérico que permite resolver el problema de segmentación de una imagen, desde un punto de vista relajado, pues permite segmentar la imagen en un número de la forma 2^N colores. Implementamos una variante del esquema del funcional de Ambrosio & Tortorelli, el cual aproxima en el sentido de la Γ-convergencia al funcional propuesto por Mumford & Shah, en esencia nuestra variante es inspirada por los trabajos de T. Chan.

Abstract

This PhD thesis, to obtain the degree of Doctor in Engineering, speciality: Applied Mathematics, is devoted to the study of an inverse problem, the mathematical analysis of a fluid-structure interaction system and the numerical implementation of a segmentation algorithm.

This work is divided in three parts. First we investigate (in Chapter 2) an inverse problem coming from Biology. We are interested in recovering the density of the ion channel (**CNG**-channel) in the olfactory cilium, from the measurement of the current flux produced by the diffusion of the agent in the cilium. This problem is described by a linear Fredholm equation of the first kind with diffusive kernel. Considering an approximation of the kernel we obtain the following results:

a) Identifiability Problem: Under different assumptions about of the regularity of the density function we establish the injectivity of the operator under study.

b) Stability Problem: We prove the stability (i.e. the continuity of the inverse of the operator) in different norms in a particular case, as well as the stability in the general case in weighted L^2-norms using Mellin's transform.

c) The Reconstruction Problem: For an appropriate choice of the parameters, we reconstruct the inverse operator.

d) Numerical Reconstruction: Using the inverse operator in the reconstruction problem, we implement a numerical algorithm to reconstruct the density.

Next, in Chapter 3, we investigate the controllability of an under-actuated underwater vehicle immersed in an infinite volume of ideal fluid with a potential flow. Taking as control input the flow of the fluid through a part of the boundary of the vehicle, we obtain a finite-dimensional system similar to Kirchhoff laws in which the control input appears through both linear terms (with time derivative) and bilinear terms. Applying the return method, we establish under certain geometric conditions a local controllability result for the position and velocities of the vehicle with four controls.

Finally, in Chapter 4, we develop a numerical algorithm allowing to solve the segmentation problem for a picture in a soft sense, limiting ourselves to numbers of colors of the form 2^N. We implement some variant of Ambrosio & Tortorelli scheme, which is an approximation in the sense of the Γ-convergence of Mumford & Shah functional. Our approach is in part inspired by T. Chan's works.

Agradecimientos

Deseo agradecer a todos los que hicieron posible la realización de esta tesis. Han transcurrido varios años de trabajo y esfuerzo, como siempre me ha dicho mi amigo, esté trabajo es la culminación y a la vez el comienzo de una etapa de mi vida. Como en todo tipo de proyectos han ocurrido cosas buenas y malas, pero siendo honesto he disfrutado este periodo de mi vida y agradezco el haber podido vivirlo.

Soledad, mi compañera, amiga y orientadora, te agradezco por tu incondicional apoyo y consejos en los momentos difíciles, te agradezco que me permitas disfrutar de tu compañía. Corina, que puedo decir de un ángel, gracias por comprender las innumerables ocasiones en las que tuviste que esperar por tú papá o simplemente comprender y entretenerte sola con cocodrilo, admiro tu dedicación y preocupación por los detalles, tienes muchos talentos y espero poder guiarte y cuidarte mucho. Bebé, no has llegado aun, pero te esperamos con alegría, no has sido parte de esta etapa, pero serás participe de lo que viene, lo que puedo decirte, es que aun cuando las cosas se ven difíciles antes de realizarlas, lo más difícil es comenzar, una vez iniciado, tarde o temprano terminara.

Quisiera dar las gracias a mis padres Marta y Gonzalo, nunca pensé en llegar a este punto en mi vida, creo que ustedes tampoco lo pensaron. Mamá, gracias por cuidarme y preocuparse por mi y mi familia, por llamar y mantener el contacto, por haberme inculcado la superación. Papá, sin duda mi héroe mi modelo, solo espero algún día poder llegar a ser la mitad de la persona que usted es. Nunca me ha molestado estar a su sombra, solo me ha inspirado, siempre he pensado que nunca hubiese podido realizar las incontables proezas que usted ha realizado a lo largo de su vida, gracias.

Agradezco también a mis profesores por su constancia y dedicación en mostrarme y enseñarme la forma de afrontar esta profesión.

A mis amigos, quisiera expresarles mi gratitud por los valores que han sembrado en mi, ademas de aprender que los sueños se pueden hacer realidad, ya que no hay cosas imposibles solo seres incapaces, Juanito, gracias por mostrarme de lo que soy capaz, extraño nuestras conversaciones, las cuales en cierto sentido me han motivado a seguir esta carrera. No tengo suficientes palabras para dar las gracias y enumerar las cosas que han hecho por mi, gracias Pablo, Jaime, Lionel, Fernando, Takeshi, Alexis, Erica, Andrés, Alejandra, Duvan, Miguel, Señores Darmois, Dugan, Ivonne, Rafa, Alejandro, Marcia, ..., y mil disculpas a quienes he olvidado mencionar, pero a todos muchas gracias.

Por último quisiera agradecer al Centro de Modelamiento Matemático CMM de la U. de Chile.

Contents

Chapter 1

Introduction

This thesis is divided in three parts.

In Chapter 2 we study a biological inverse problem and we focus on a linearization of this problem, considering the following issues:

a) Identifiability problem.

b) Stability problem.

c) Reconstruction problem.

d) Numerical algorithm for reconstruction.

Next, in Chapter 3, we study a control problem coming from fluid-structure interaction systems. More precisely, we investigate the controllability of the motion of a submarine, which is surrounded by an inviscid incompressible fluid.

Finally, in Chapter 4, we develop a numerical algorithm to obtain the segmentation of a picture in the soft sense, as we consider a finite number of colors.

Now we present a short description of the problems and the main results obtained in this thesis.

1.1 Inverse problem in biology

Identification of detailed features of neuronal systems is an important challenge in the biosciences today. Olfactory cilia are thin hair-like filaments that extend from olfactory receptor neurons into the nasal mucus. Transduction of an odor into an electrical signal is accomplished by a depolarizing influx of ions through cyclic-nucleotide-gated (CNG) channels in the membrane that forms the lateral surface of the cilium and are activated by adenosine 3', 5'-cyclic monophosphate (cAMP). In an experimental procedure developed by S. Kleene, a cilium is detached at its base and drawn into a recording pipette. The cilium base is then immersed in a bath of a channel activating agent (cAMP) which is allowed to diffuse into the cilium interior, opening channels as it goes and initiating a transmembrane current. The total current is recorded as a function of time.

In [37], French et. al. propose a mathematical model consisting of two nonlinear differential equations and a constrained Fredholm integral equation of the first kind is developed to model this experiment. The unknowns in the problem are the concentration c of cAMP, the membrane potential V and, the quantity of most interest in this work, the distribution of CNG channels ρ along the length of a cilium. Then the basic idea is to recover the distribution of CNG channels along the length of a cilium only measuring the electrical activity produced by the diffusion of cAMP into cilia, this constitute an interesting inverse problem.

French et. al. in [37] present a simple numerical method to obtain estimates of the spatial distribution of CNG ion channels along the length of a cilium. Certain computations indicate that this mathematical problem is ill-conditioned.

Later, French and Edwards in [36], studied this inverse problem using perturbation techniques. A simple perturbation approximation is derived and used to solve this inverse problem and thus obtain estimates of the spatial distribution of CNG ion channels along the length of a cilium. A one-dimensional computer minimization and a special delay iteration are used with the perturbation formulas to obtain approximate channel distributions in the cases of simulated and experimental data.

French and Groetsch in [38] present some simplifications and approximations in the problem, obtaining an analytical solution for this inverse problem and a numerical procedure is proposed for a class of integral equations suggested by this simplified model and numerical results using simulated and laboratory data are presented.

In this work we consider the linear problem proposed in [38], but with a better approximation of the operator kernel and we study the identifiability, stability and numerical reconstruction for this inverse problem.

1.1.1 Linear problem

Now we introduce the lineal model presented in [38]. In [37] a nonlinear integral equation model is developed for determining the spatial distribution of ion channels along the length of frog olfactory cilia. The essential nonlinearity in the model arises from the binding of the channel activating ligand to the cyclic-nucleotide-gated ion channels as the ligand diffuses along the length of the cilium. We investigate a linear model of this process in which the binding mechanism is neglected, leading to a particular type of linear Fredholm integral equation of the first kind with a diffusive kernel. The linear mathematical model consists of finding $\rho = \rho(x) > 0$, such that

$$I[\rho](t) = J_0 \int_0^L \rho(x) K(t,x) dx, \tag{1.1.1}$$

where the kernel is

$$K(t,x) = F(c(t,x)), \tag{1.1.2}$$

and c is the concentration the of the problem

$$\begin{cases} \dfrac{\partial c}{\partial t} - D\dfrac{\partial^2 c}{\partial x^2} &= 0, \quad t>0,\ x\in(0,L) \\ c(0,x) &= 0, \qquad x\in(0,L) \\ c(t,0) &= c_0, \qquad t>0 \\ \dfrac{\partial c}{\partial x}(t,L) &= 0, \qquad t>0, \end{cases} \tag{1.1.3}$$

with

$$F(x) = \frac{x^n}{x^n + K_{1/2}^n}. \tag{1.1.4}$$

$F(x)$ is a "Hill" function with constant $K_{1/2}$ and exponent n, where $K_{1/2}$ is the half-bulk concentration and n is an experimentally determined parameter. The function ρ is an unknown ion channel density function, c is the concentration of a channel activating ligand that is diffusing from left-to-right in a thin cylinder (the interior of the cilium) of length L with diffusivity constant D, $I[\rho](t)$ is a given total transmembrane current, the constant J_0 has units of current/length, and c_0 is the maintained concentration at the open end of the cylinder (while $x = L$ is considered the closed end).

This problem arises in models of diffusion in tiny tubular biological structures. In physical experiments a ligand (cAMP), which is held at constant concentration c_0 at the open end, diffuses into the cilia and binds to ion channels (CNG) opening the channels and thereby allowing an ionic influx which initiates a local transmembrane current. The model (1.1.1)-(1.1.3) does not include terms modeling the binding or changes in membrane potential. However, if the quantity of ion channels is small these effects will be negligible. Frog olfactory cilia have diameters of about 0.28μm and length around 40μm and are connected to the knobs at the ends of the dendrites of the olfactory receptor neurons. The Hill function in the experiments modeled here typically has exponent $n = 1.7$ and half-concentration $K_{1/2} = 1.7\mu$m. Here $J_0 = 0.232$ pA/channel is a positive constant.

Now we derive an analytical solution of a simplified version of problem (1.1.1)-(1.1.3). To that end we assume:

i) The influence of the no-flux boundary condition for c at the closed end is negligible (or equivalently that the cilium is very long). Neglecting the boundary condition at the closed end, the diffusion equation has the closed form solution

$$c(t,x) = c_0 \,\mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right), \tag{1.1.5}$$

where erfc is the complementary error function defined by

$$\mathrm{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}}\int_0^z \exp(-\tau^2)d\tau. \tag{1.1.6}$$

Therefore, curves of the form $t \propto x^2$ are level curves for the concentration function c.

ii) The Hill function exponent n is large.

This assumption leads to the following simplification;

$$\begin{aligned} F(c(t,x)) &= \frac{c(t,x)^n}{c(t,x)^n + K_{1/2}^n} && (1.1.7)\\ &\simeq H(c(t,x) - K_{1/2}) && (1.1.8)\\ &= H(\beta^2 t - x^2) && (1.1.9)\end{aligned}$$

where H is the Heaviside unit step function

$$H(u) = \begin{cases} 0 & \text{if } u < 0, \\ 1/2 & \text{if } u = 0, \\ 1 & \text{if } u > 0, \end{cases} \tag{1.1.10}$$

and $\beta \sim \sqrt{D}$ is implicitly given in terms of the half bulk concentration by the equation

$$K_{1/2} = c_0 \mathrm{erfc}(\beta/(2\sqrt{Dt})). \tag{1.1.11}$$

This type of assumption has also been made in the study of calcium sparks (see Section 8.11 by [28]).

Using these assumptions we have

$$I[\rho](t) = J_0 \int_0^L H(\beta^2 t - x^2)\rho(x)dx = J_0 \int_0^{\beta\sqrt{t}} \rho(x)dx.$$

Differentiating yields

$$I'(t) = \frac{1}{2}\beta t^{-1/2} J_0 \rho(\beta\sqrt{t}).$$

Letting $y = \beta\sqrt{t}$, we obtain

$$\rho(y) = \frac{2I'(y^2/\beta^2)y}{J_0\beta^2}.$$

In this simplified model, the channel density is given analytically in terms of the derivative of the total current and a parameter β which is obtained by solving equation (1.1.11) numerically.

Our idea to consider a more general situation. We replace (1.1.8) by the following approximation of the Hill function (1.1.4)

$$F(x) \simeq F_m(x) = F(c_0)\sum_{j=1}^{m} a_j H(x - \alpha_j), \quad \forall x \in [0, c_0], \tag{1.1.12}$$

where a_j, α_j are positive constants verifying:

$$\sum_{j=1}^{m} a_j = 1, \tag{1.1.13}$$

and

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m < c_0, \tag{1.1.14}$$

thus $\{\alpha_j\}_{j=1}^m$ define a mesh on the interval $(0, c_0)$. With the above assumptions we define the approximate total current

$$I_m[\rho](t) = J_0 \int_0^L \rho(x) K_m(t,x) dx, \tag{1.1.15}$$

where

$$K_m(t.x) = F_m(c(t,x)) = F_m\left(c_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)\right). \tag{1.1.16}$$

Therefore, our inverse problem is to recover ρ from the measurement $I_m[\rho](t)$, for all $t > 0$.

1.1.2 Main results

Now we present the main results in this work. We begin by setting the notations. For any $\gamma > 0$, we consider the weight function $\sigma_\gamma(x) = |x|^\gamma$, and introduce the following norms

$$\begin{aligned}
\|f\|_{0,\gamma,b} &= \|\sigma_\gamma f\|_{L^2(0,b)}, \\
\|f\|_{1,\gamma,b} &= \|\sigma_\gamma f\|_{H^1(0,b)}, \\
\|f\|_{-1,\gamma,b} &= \|\sigma_\gamma f\|_{H^{-1}(0,b)},
\end{aligned}$$

and we set the constants

$$L_k = L/\beta_k, \quad \text{for } k = 1, ..., m, \quad \text{and } L_0 = 0, \tag{1.1.17}$$

where

$$\beta_j = \operatorname{erfc}^{-1}(\alpha_j/c_0) 2\sqrt{D} \ \text{ for } j = 1, ..., m. \tag{1.1.18}$$

First, we present a result about the continuity of the operator I_m.

Theorem 1.1.1. *Let $\rho : [0, L] \to \mathbb{R}$ be a function in $L^2(0, L)$. Then for $\gamma \geq \frac{3}{4}$, there exists a positive constant $C_1 > 0$, such that,*

$$\|I_m[\rho]\|_{1,\gamma,L_m^2} \leq C_1 \|\rho\|_{L^2(0,L)}, \tag{1.1.19}$$

where C_1, only depends on $L, \alpha_1, \alpha_{m-1}, \alpha_m, a_m$ and γ.

The above theorem sets the functional framework for I_m. Now we present our main results of the stability. First we define the constant

$$\gamma_0 = \max\left\{0, \frac{\ln(a_m)}{\ln\left(\frac{\beta_m}{\beta_{m-1}}\right)} - \frac{1}{2}\right\}. \tag{1.1.20}$$

Theorem 1.1.2. *Let $\rho : [0, L] \to \mathbb{R}$ be a function in $L^2(0, L)$. Then, for all $\gamma > \max\{\gamma_0, 3/4\}$, there exists a positive constant $C_2 > 0$, such that,*

$$\|\rho\|_{-1,\gamma+1,L} \leq C_2 \|I_m[\rho]\|_{1,\frac{\gamma}{2}-\frac{1}{4},L_m^2}, \tag{1.1.21}$$

where C_2, only depends on $L, \alpha_1, \alpha_{m-1}, \alpha_m, a_m$ and γ.

This theorem corresponds to the observability inequality for the functional I_m, in particular, this result proves the identifiability for our inverse problem. For the proof see Section 2.2.

In the proof of the above result, we can see the role played by the mesh $\{\alpha_j\}_{j=1}^m$, for $j = 1, ..., m-1$. We will focus our analysis in the special case where

$$\alpha_j = c_0 \text{ erfc}\left(\frac{\beta_0\beta^j}{2\sqrt{D}}\right) \quad j = 1, ..., m, \tag{1.1.22}$$

With $\beta \in (0, 1)$ and $\beta_0 > 0$ constants. This simplification is not very restrictive because $\{\alpha_j\}_{j=1}^m$ defined a mesh in the interval $[0, c_0]$, therefore (1.1.22) only fixed a non-regular mesh in which we want to have the approximation (1.1.12), but this choice allows us to obtain a reconstruction algorithm for the density ρ.

Firstly, we study the following functional

$$\Phi_m[\varphi](t) = \frac{1}{J_0F(c_0)} I_m[\rho](t^2) \quad \forall t > 0, \tag{1.1.23}$$

where

$$\varphi(x) = \int_0^x \rho(z)dz. \tag{1.1.24}$$

It is worth noting that the identifiability for Φ_m is equivalent to the identifiability for I_m, for that reason we will focus in the study of identifiability on I_m.

In what follows $\tilde{I}_m$ and $\tilde{\Phi}_m$ the functionals I_m, Φ_m respectively, the choice of α_j as in (1.1.22).

Thus, for $\rho \in \mathcal{C}^0([0, L])$, we can construct a function $\tilde{\varphi} : [0, L] \to \mathbb{R}$ depending only on ρ, by the equation define

$$\int_0^x \rho(z)dz = \tilde{\varphi}(x) + \frac{\tilde{I}_m[\rho](L_m^2)}{J_0F(c_0)} \quad \forall x \in [0, L], \tag{1.1.25}$$

A: approximation of ρ. **B**: current $I(t)$ as defined by (1.1.26)

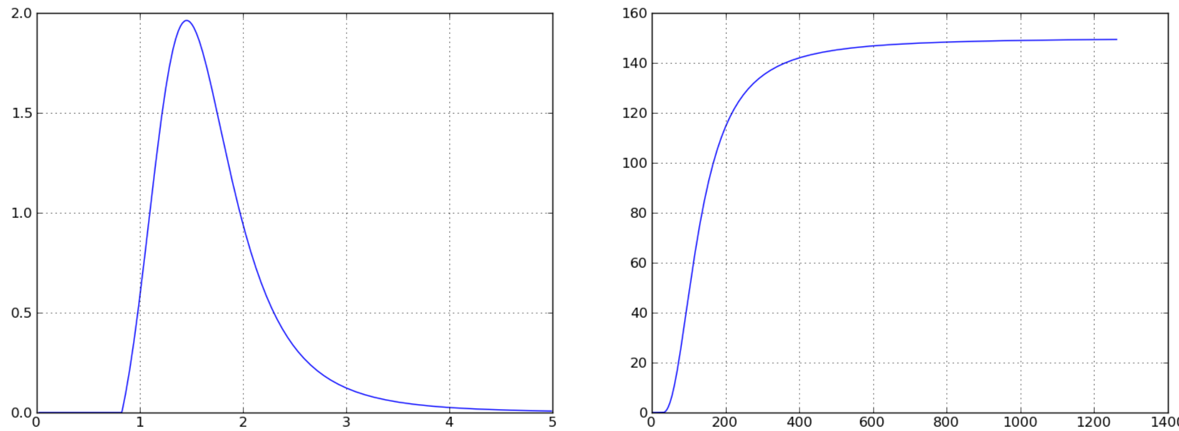


Figure 1.1: Approximation of the function $\rho(x)$ and with current $I(t)$ as defined by (1.1.26)

therefore, deriving (1.1.25) we can recover ρ (see Theorem 2.2.9). Moreover the equation (1.1.25) define a numerical algorithm for the reconstruction problem (see Section 2.9 on page 59). If we consider the same example considered in [38], thus we define

$$I(t) = \begin{cases} 0, & t \in (0, t_{Delay}), \\ I_{\text{Max}} \left[1 + \left(\dfrac{K_I}{t - t_{Delay}} \right)^{n_I} \right]^{-1}, & t > t_{Delay}, \end{cases} \tag{1.1.26}$$

with $t_{Delay} = 30ms$, $n_I = 2.2$, $I_{Max} = 150pA$ and $K_I = 100ms$. The current (1.1.26) has a short delay (Figure 2.5**B**), which is similar to the profiles in several applications (see [18], [33] or [58]).

Thus we obtain the numerical solution as shown in Figure 1.1**A**. In this figure one sees that the main features of the solution of this model are consistent with those obtained in [33, 38].

With our approach is possible to consider a better and more realistic approximation of the mathematical model (1.1.1). Furthermore we have developed a reconstruction method and not only a numerical approach of the model. On other hand, we have obtained a stability result with different norms.

1.1.3 Polynomial approximation

We study a second approximation for the kernel (1.1.2), in this case, we consider a polynomial approximation of the Hill's function (1.1.4), around c_0.

We consider P_m, the Taylor expansion of degree m. For Hill's function defined by (1.1.4) at point $c_0 > 0$. Thus we have

$$F(x) = P_m(x - c_0) + O(|x - c_0|^{m+1}). \tag{1.1.27}$$

Therefore, we define the polynomial kernel approximation by

$$PK_m(t.x) = P_m(w(t, x) - c_0), \tag{1.1.28}$$

where $w(t, x)$ is the solution of the problem (1.1.3).

The total current with polynomial approximation is given by

$$PI_m[\rho](t) = \int_0^L \rho(x) PK_m(t, x)dx, \quad \forall t > 0. \tag{1.1.29}$$

Now we present our principal result related with the above for $\rho \in L^2(0, L)$.

Theorem 1.1.3. *Let $m \leq 8$, be a constant. Then we have*

$$Ker\ PI_m = \{0\},$$

where

$$Ker\ PI_m = \left\{ f \in L^2(0, L) \middle| \ PI_m[f](t) = 0,\ \forall t > 0 \right\}.$$

Theorem 1.1.3 corresponds to the identifiability of the operator PI_m. The proof is given in Section 2.10.

1.1.4 A non-linear inverse problem

We consider a variation of the olfactory inverse problem introduced by French et al. in [37] and we will show that, this inverse problem corresponds to determine a potential q in an interval of $\mathbb{R}$, from the Neumann data measured on one side of the boundary for the equation $-\Delta + q$. This is correlated with the problem of determining a complex-valued potential q in a bounded two dimensional domain, from the Cauchy data measured on an arbitrary open subset of the boundary for the associated Schrodinger equation $-\Delta + q$, the identifiability for this problem was shown in [48]. The motivation for this problem comes from a classical inverse problem of electrical impedance tomography. In this case one attempts to determine the electrical conductivity of a body, considering measurements of voltage and current on the boundary of the body. This problem was proposed by A. Calderón [14] and is well known as *Calderón's problem* and had been study by several authors. We proof the non-identifiability for our problem, showing the importance of result [48].

Let us define $q(x) = \mathbb{P}R_a F(c_0)\,\rho(x)$. We consider the stationary case, i.e. when the functions do not depend on time in the model develop in [37]. Thus, we can reduce this inverse problem to the following form: We want to determine the positive potential $q(x)$ such that:

$$\begin{cases} v'' - qv = 0, & x \in (0, L) \\ v(0) = -v_0, & \\ v'(L) = 0, & \end{cases} \tag{1.1.30}$$

the measurement $v'(0)$, where $q(x) = \mathbb{P}R_a F(c_0)\,\rho(x)$ this allow us to obtain ρ.

Our result corresponds to a non-identifiablity result.

Theorem 1.1.4. (Non-identifiability) *Given a constant $v_0 > 0$, there exist two different positives potential $q_1, q_2 \in C^{1+\alpha}([0, L])$, for some $\alpha > 0$, such that the respective solutions of (1.1.30), v_1, v_2, satisfy*

$$v_1'(0) = v_2'(0). \tag{1.1.31}$$

Despite of this non-identifiability result, it helps us to understand the original identifiability problem and gives us an insight about the difficulties of the problem. For the proof see Section 2.11.

1.2 Control of underwater vehicles in potential fluids

In this work, we investigate the controllability of an underactuated underwater vehicle immersed in an infinite volume of ideal fluid with a potential flow. Taking as control input the flow of the fluid through a part of the boundary of the vehicle, we obtain a finite-dimensional system similar to Kirchhoff laws in which the control input appears through both linear terms (with time derivative) and bilinear terms. Applying the return method, we establish under certain geometric conditions a local controllability result for the position and velocities of the vehicle with four controls. We prove that the position, orientation, and velocity of the vehicle are locally controllable with four control input.

The control of boats or submarines has attracted the attention of the mathematical community from a long time, see for instant the works by [60, 61, 62, 15]. In most of the papers devoted to that issue, the fluid is assumed to be inviscid, incompressible and irrotational, and the rigid body is supposed to have an elliptic shape. On the other hand, to simplify the model, the control is often assumed to appear in a linear way in a finite-dimensional system describing the dynamics of the rigid body, the so-called Kirchhoff laws. A large vessel (e.g. a freighter) presents often one tunnel thruster built into the bow to make docking easier.

Glass and Rosier in [44] studied a problem in $2D$, i.e. $S \subset \mathbb{R}^2$ with one axis of symmetry, surrounded by a fluid, and controlled by two fluid flows, a longitudinal one and a transversal one (see Figure 1.2).

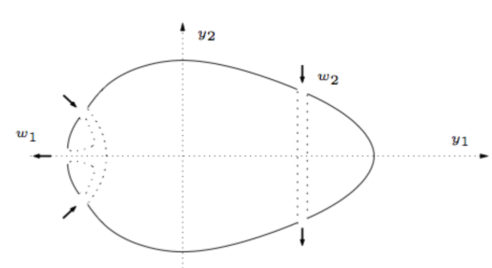


Figure 1.2: Forward propulsion by a propeller

They aimed to control the position and velocity of the rigid body by the control inputs. For this system, the state lives in $\mathbb{R}^6$ (but there is a PDE in the dynamics) and the control inputs lives in $\mathbb{R}^2$ and there is no control objective for the fluid flow, as it is an exterior domain. It provides an accurate model for the motion of a boat with two propellers, one displayed in a transversal bow thruster at the bow of the ship, the other one placed at the stern of the boat (see Figure 1.3). A rigorous analysis of the control properties of such a system is performed in [44].

Bow thruster

Longitudinal thruster

Figure 1.3: Example of the a bow thruster and longitudinal thruster

In this thesis we aim to extend this work to the dimension three. We investigate the motion of under-actuated underwater vehicles (for instance submarines).

The fluid, still inviscid and incompressible, will no longer be assumed to be irrotational (i.e., the vorticity may not vanish everywhere), and the only geometric assumption for the shape of the submarine will be the existence of two planar symmetries.

To be more precise, our fluid-structure interaction problem can be described as follow. The underwater vehicle, represented by a rigid body occupying a connected compact set $S(t) \subset \mathbb{R}^3$, is surrounded by an homogeneous incompressible perfect fluid filling the open set $\Omega(t) := \mathbb{R}^3 \setminus S(t)$ (as for e.g. a submarine immersed in an ocean). We assume that $\Omega(t)$ is C^∞ smooth and connected. Let $S = S(0)$ and

$$\Omega = \Omega(0) = \mathbb{R}^3 \setminus S(0)$$

denote the initial configuration ($t = 0$). Then, the dynamics of the fluid-structure system is governed by the following system of PDE's

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p = 0, \qquad t \in (0,T),\ x \in \Omega(t) \tag{1.2.1}$$

$$\text{div } u = 0, \qquad t \in (0,T),\ x \in \Omega(t) \tag{1.2.2}$$

$$u \cdot \hat{n} = (h' + \omega \times (x - h)) \cdot \hat{n} + w(t,x), \qquad t \in (0,T),\ x \in \partial\Omega(t) \tag{1.2.3}$$

$$\lim_{|x| \to +\infty} u(t,x) = 0, \tag{1.2.4}$$

$$mh'' = \int_{\partial\Omega(t)} p\hat{n}\, d\sigma, \qquad t \in (0,T) \tag{1.2.5}$$

$$\frac{d}{dt}(QJ_0Q^*\omega) = \int_{\partial\Omega(t)} (x - h) \times p\hat{n}\, d\sigma, \qquad t \in (0,T) \tag{1.2.6}$$

$$Q' = S(\omega)Q, \qquad t \in (0,T) \tag{1.2.7}$$

$$u(0,x) = u_0(x), \qquad x \in \Omega \tag{1.2.8}$$

$$(h(0), Q(0), h'(0), \omega(0)) = (h_0, Q_0, h_1, \omega_0) \quad \in \mathbb{R}^3 \times \mathrm{SO}(3) \times \mathbb{R}^3 \times \mathbb{R}^3. \tag{1.2.9}$$

In the above equations, u (resp. p) is the velocity field (resp. the pressure) of the fluid, h denotes the position of the center of mass of the solid, ω denotes the angular velocity and Q the orthogonal matrix giving the orientation of the solid. The positive constant m and the matrix J_0, which denote respectively the mass and the inertia matrix of the rigid body, are defined as

$$m = \int_S \rho(x)dx, \quad J_0 = \int_S \rho(x)(|x|^2 Id - xx^*)dx,$$

where $\rho(\cdot)$ represents the density of the rigid body. $\hat{n}$ is the outward unit vector to $\partial\Omega(t)$, $x \times y$ is the cross product between the vectors x and y, and $S(y)$ is the skew-adjoint matrix such that $S(y)x = y \times x$, i.e.

$$S(y) = \begin{pmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{pmatrix}.$$

Finally, the term $w(t,x)$, which stands for the flow through the boundary of the rigid body, is taken as control input. Its support will be strictly included in $\partial\Omega(t)$, and actually only a finite dimensional control input will be considered here (see below (3.2.16) for the precise form of the control term $w(t,x)$).

Here, we are interested in the control properties of (1.2.1-1.2.9). The controllability of Euler equations has been established in 2D (resp. in 3D) in [23] (resp. in [42]). Note, however, that there is no hope here to control both the fluid and the rigid body motion. Indeed, $\Omega(t)$ is an exterior domain, and the vorticity is transported by the flow with a finite speed propagation, so that it is not affected (at any given time) far from the boat. Therefore, we will deal with the control of the motion of the rigid body only. As the state of the rigid body is described by a point in $\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathbb{R}^3 \times \mathbb{R}^3$, it is natural to consider a finite-dimensional control input.

Consider the matrix $\mathcal{J} \in \mathbb{R}^{6\times 6}$ introduced in 3.3.31, which represents a generalization of the mass and the inertia matrix for the fluid-struct system (1.2.1-1.2.9).

Let $(l, r) \in \mathbb{R}^3 \times \mathbb{R}^3$ denote the part of linear and angular velocities of the body and introduce the corresponding pulses $(P, \Pi) \in \mathbb{R}^3 \times \mathbb{R}^3$ defined by

$$\mathcal{J}\begin{pmatrix} l \\ r \end{pmatrix} = \begin{pmatrix} P \\ \Pi \end{pmatrix}. \tag{1.2.10}$$

Then Proposition 3.3.2 provides the following equations for the dynamics of the rigid body

$$\begin{aligned} \frac{dP}{dt} + C^M \dot{w} &= (P + C^M w) \times r - \sum_{1\leq p\leq n} w_p \left\{ L_p^M l + R_p^M r + W_p^M w \right\}, \\ \frac{d\Pi}{dt} + C^J \dot{w} &= (\Pi + C^J w) \times r + P \times l - \sum_{1\leq p\leq n} w_p \left\{ L_p^J l + R_p^J r + W_p^J w \right\}, \end{aligned} \tag{1.2.11}$$

where $w(t) := (w_1(t), ..., w_n(t)) \in \mathbb{R}^n$ denotes the control input and the matrixes C^M, C^J, L_p^M, R_p^M, W_p^M, L_p^J, R_p^J y W_p^J only depend on the geometry of the body and the position of the controls on the body surface.

If we linearize the above system, we obtain

$$\begin{pmatrix} l' \\ r' \end{pmatrix} = -\mathcal{J}^{-1} \begin{pmatrix} C^M \\ C^J \end{pmatrix} w', \tag{1.2.12}$$

which is controllable if the rank of the matrix

$$\begin{pmatrix} C^M \\ C^J \end{pmatrix},$$

is six. Therefore, we need to consider at least six controls input.

The work is organized as follows. In Section 3.3, we simplify system (1.2.1-1.2.9) by assuming that the fluid is potential. We obtain a finite dimensional system (namely (3.3.64)) similar to Kirchhoff laws (1.2.11), in which the control input w appears through both linear terms (with time derivative) and bilinear terms. The investigation of the control properties of (3.3.64) is performed in Section 3.4. After noticing that the controllability of the linearized system at the origin requires six control inputs, we apply the *return method* due to Jean-Michel Coron to take advantage of the nonlinear terms in (3.3.64). (We refer the reader to [24] for an exposition of that method for finite-dimensional systems and for PDEs.) We consider the linearization along a certain closed-loop trajectory and obtain a local controllability result (Theorem 3.4.9) assuming that two rank conditions are fulfilled, by using a variant of Silverman-Meadows test for the controllability of a time-varying linear system. Some examples using symmetry properties of the rigid body are given in Section 3.5.

1.3 Image problem

This section is concerned with image segmentation, which plays a very important role in many applications. The aim is to find a partition of an image into its constituent parts. As we will see, the main difficulty is that one needs to manipulate objects of different kinds: Functions, domains in $\mathbb{R}^2$ and curves. We could say that segmenting an image means dividing it into its constituent parts.

Our research aims at image segmentation using the variational framework of Mumford and Shah, following an approximation proposed by Ambrosio and Tortorelli. This technique circumvents the use of parametric contours and implicit level-set techniques, where its solution may be regarded as a soft segmentation, with a number of levels or colors being 2^N. On the other hand, the implementation was based on an finite difference discretization, where two- and four- color cases are described with their corresponding numerical results. See Chapter 4.

1.3.1 The Mumford and Shah functional

Let us present the model introduced by Mumford and Shah in 1989 [70]. In this section Ω is a bounded open set of $\mathbb{R}^N$, $N = 2, 3$, and $u_0(x)$ is the initial image. Without loss of generality we can always assume that $0 \le u_0(x) \le 1$ a.e. $x \in \Omega$. We search for a pair (u, K), where $K \subset \Omega$ is the set of discontinuities, minimizing

$$F(u,K) = \int_{\Omega\setminus K} (u - u_0)^2 dx + \alpha \int_{\Omega\setminus K} |\nabla u|^2 dx + \beta \mathcal{H}^{N-1}(K),$$

where α and β are nonnegative constants and $\mathcal{H}^{N-1}(K)$ corresponds to $N-1$ dimensional Hausdorff measure of K, which is the most natural way of extending the notion of length to non-smooth sets and u belongs to the Sobolev space $W^{1,2}(\Omega \setminus K)$.

The difficulty in studying F is that it involves two unknowns u and K of different natures: u is a function defined on an N-dimensional space, while K is an $(N-1)$-dimensional set.

In order to apply the direct method of the calculus of variations, it is necessary to find a topology that ensures at the same time lower semicontinuity of F and compactness of the minimizing sequences. The difficulty comes from $H^{N-1}(K)$. Indeed, let E be a Borel set of $\mathbb{R}^N$ with topological boundary ∂E. It is easy to convince oneself that the map $E \to H^{N-1}(E)$. is not lower semicontinuous with respect to any compact topology.

This shows the necessity of finding another formulation of $F(u, K)$. The new formulation involves the space $BV(\Omega)$ of functions of bounded variation in Ω. The idea is to identify the set of edges K with the jump set S_u of u, which allows us to eliminate the unknown K. So the idea is to consider the functional

$$G(u) = \int_{\Omega} (u - u_0)^2 dx + \alpha \int_{\Omega} |\nabla u|^2 dx + \beta \mathcal{H}^{N-1}(S_u).$$

Now, it is tempting to minimize G on the space $BV(\Omega)$. Unfortunately, the space $BV(\Omega)$ may contain pathological non-constant functions that are continuous and have approximate differential equal to zero almost everywhere (a well-known example is the Cantor-Vitali function). For such a function v we have

$$G(v) = \int_\Omega (v - u_0)^2 dx \geq \inf_{u \in BV(\Omega)} G(u),$$

and since these pathological functions are dense in $L^2(\Omega)$, we get

$$\inf_{u \in BV(\Omega)} G(u) = 0,$$

which implies that the infimum of G cannot be achieved in $BV(\Omega)$ in general.

To avoid this phenomenon we must eliminate these pathological functions, which have the peculiarity that their distributional derivatives are measures concentrated on Cantor sets. Let us recall that the distributional derivative Du of a $BV(\Omega)$ function can be split into three mutually singular measures:

$$Du = \nabla u dx + (u^+ - u^-) n_u \mathcal{H}^{N-1}_{|S_u} + C_u,$$

where $J(u) = (u^+ - u^-) n_u \mathcal{H}^{N-1}_{|S_u}$ is the jump part and C_u the Cantor part. Following Di Giorgi [40, 39] we define $SBV(\Omega)$ as the space of special functions of bounded variation, which is the space of $BV(\Omega)$ functions such that $C_u = 0$. The natural question is now to establish the relation between the following two problems:

$$\inf_{u,K} \left\{ \begin{array}{ll} F(u,K), & u \in W^{1,2}(\Omega \setminus K) \cap L^\infty(\Omega) \\ & K \subset \Omega, \ \ K \text{ closed}, \ \ \mathcal{H}^{N-1} < \infty \end{array} \right\}, \tag{$P1$}$$

$$\inf_u \left\{ \ G(u), \ \ u \in SBV(\Omega) \cap L^\infty(\Omega) \ \right\}. \tag{$P2$}$$

The answer can be found in Ambrosio [3] and is the consequence of the following theorem:

Theorem 1.3.1. *[3] Let $K \subset \Omega$ be a closed set such that $\mathcal{H}^{N-1}(K) < \infty$ and let $u \in W^{1,2}(\Omega \setminus K) \cap L^\infty(\Omega)$. Then $u \in SBV(\Omega)$ and $S_u \subset K \cup L$ with $\mathcal{H}^{N-1}(L) = 0$.*

From Theorem 1.3.1 it follows that $\inf P2 \leq \inf P1$. By using compactness and lower semicontinuity theorems (see below) it can be shown that $(P2)$ has a solution u. For such a minimizer De Giorgi-Carriero-Leaci [41] proved that

$$\mathcal{H}^{N-1}(\Omega \cap (\overline{S_u} \setminus S_u)) = 0.$$

So by setting $K = \Omega \cap \overline{S_u}$ we get a solution of $(P1)$ and $\min(P1) = \min(P2)$. It remains to show that $(P2)$ has a solution. This is a direct consequence of the following theorem:

Theorem 1.3.2. *[4] Let $u_n \in SBV(\Omega)$ be a sequence of functions such that there exists a constant C with $|u_n(x)| \leq C < \infty$ for a.e. $x \in \Omega$ and $\int_\Omega |\nabla u_n|^2 dx + \mathcal{H}^{N-1}(S_{u_n}) \leq C$. Then there exists a subsequence u_{n_k} converging a.e. x to a function $u \in SBV(\Omega)$. Moreover, ∇u_{n_k} converges weakly in $L^2(\Omega)^N$ to ∇u, and $\liminf \mathcal{H}^{N-1}(S_{u_{n_k}}) \geq \mathcal{H}^{N-1}(S_u)$.*

We obtain a solution for $(P2)$ by applying Theorem 1.3.2 to any minimizing sequence of $(P2)$ and by observing beforehand that we can restrict our attention to minimizing sequences satisfying $|u_n|_{L^\infty(\Omega)} \leq |u_0|_{L^\infty(\Omega)}$ (using a truncation argument).

1.3.2 Approximations of the Mumford and Shah functional

The lack of differentiability of the functional for a suitable norm does not allow us to use, as is classical, Euler-Lagrange equations. Moreover, the discretization of the unknown discontinuity set is a very complex task. A commonly used method is to approximate $F(u, K)$ (or $G(u)$) by a sequence F_ε of regular functionals defined on Sobolev spaces, the convergence of F_ε to F as $\varepsilon \to 0$ being understood in the $\Gamma-$convergence framework. Of course, if we want to get an efficient approximation, the set K must not appear in F_ε. Approximations by elliptic functionals were introduced in [5]. In this approach the set S_u (or K) is replaced by an auxiliary variable v (a function) that approximates the characteristic function $(1 - \chi_{S_u})$, i.e., $v(x) \simeq 0$ if $x \in S_u$ and $v(x) \simeq 1$ otherwise. Ambrosio and Tortorelli [5] proposed the following sequence of functionals:

$$F_\varepsilon(u, v) = \int_\Omega |u - u_0|^2 dx + \int_\Omega v^2 |\nabla u|^2 dx + \varepsilon \int_\Omega |\nabla v|^2 dx + \frac{1}{4\varepsilon} \int_\Omega (v - 1)^2 dx.$$

Chapter 2

Inverse problem in biology

The understanding of transduction in the olfactory system is an important challenge in neuroscience. It is worth noting that the olfactory receptor neuron is bipolar, the one end extending dendrites to the surface of the neuroepithelium and the other end sending axons that terminate in the olfactory bulb in the brain. The dendrites form knobs from which specialized cilia emanate, the transduction begins when an odorant molecule (odor) binds to receptors on the cytoplasmic membrane of a cilium, of an olfactory neuron (for reviews of olfactory transduction see [79, 11, 81]).

Each olfactory neuron has several of these thin cilia into the nasal mucus. When the odorant binds with the receptor, this causes and increased of the level of Cyclic adenosine monophosphate (cAMP) inside the cilium. cAMP is one of the intracellular messengers involved in olfactory signal transduction. This cAMP is synthesized by an adenylate cyclase, and hydrolyzed by a phosphodiesterase. This is achieved through a GTP-binding protein, see [77, 86, 85, 13, 10]. The cAMP opens cyclic nucleotide-gated (CNG) channels, the CNG channels allow a depolarizing the cell through a influx of Na^+ and Ca^{2+} into the cilia [31, 32].

Although the single-channel properties have been well studied and described, less is known about the distribution of these channels along the cilia. Actually it is in general unknown, and it is therefore assumed to by uniform at first approximation, even if its variations could significantly have an influence on the sensitivity of the neuron to odor stimuli. It is important to note that Ca^{2+} has a main role in olfactory transduction [81], opening Ca^{2+}-activated Cl^- channels (ClCa), that allows Cl^- flux from the cilia. This flux is one of those most responsible for depolarizing the cell. This generates an excitatory response, characterized by an increase in the frequency of action potential [53, 59, 64].

Experiments can be performed (see [18]) on single cilia; recordings are made of current as cAMP diffuses and activates the CNG channels. In this work, we intend to see whether CNG channel distributions can be derived from the experimental current data and hence if we can recover some properties of the cilia. Therefore, this problem can be represented by a mathematical model, and CNG channel distributions can be computed, in an approximate way, from the current data, considering suitable approximations (see [37]). The model involves a system of nonlinear equations that corresponds to an inverse problem in Partial Differential Equations (PDE). The perturbation theory allows to find an approximate solution (see

[36]), and in [38], under certain approximations and simplifications, was derived an integral equation model.

2.1 Setting the problem

In this section we will set the mathematical model related to the inverse problem arising in olfaction experimentation.

In order to simplify our problem, we will consider the domain $\Omega \subset \mathbb{R}^3$ as a cylinder, closed at one end and open at the other end. Note that the boundary of the cylinder corresponds to the plasmatic membrane. Thanks to the symmetry of the domain, the dynamics are described by a one dimensional model for the olfactory cilium, and we consider a diffusion-reaction model for the cAMP molecules, immersed in a solution. We denote by $c(t,x)$ the (volume) concentration of cAMP molecules. Those molecules diffuse in the x-direction in a one dimensional channel of length L. We have the Neumann boundary condition $\frac{\partial c}{\partial x}(t,L) = 0$ for $t > 0$. Moreover, in our situation the concentration of cAMP is constant at the other side, that is, $c(t,0) = c_0 > 0$ for $t > 0$, and initially the channel does not have cAMP.

Therefore, the conservation of mass is given by

$$\begin{cases} \dfrac{\partial c}{\partial t} - D\dfrac{\partial^2 c}{\partial x^2} &= 0, \quad t > 0,\ x \in (0,L) \\ c(0,x) &= 0, \quad x \in (0,L) \\ c(t,0) &= c_0, \quad t > 0 \\ \dfrac{\partial c}{\partial x}(t,L) &= 0, \quad t > 0, \end{cases} \tag{2.1.1}$$

where the constant $D > 0$ is the diffusion coefficient.

The binding of the cAMP on the cilium surface produces a change in the current flux ($J(t,x)$) through the channels. This fact is modeled by

$$J(t,x) = J_0 \rho(x) F(c(t,x)), \quad t > 0,\ x \in (0,L), \tag{2.1.2}$$

where $\rho(x)$ is the density of channels along the surface of the cilium and F is the classical dimensionless Hill function, given by

$$F(x) = \frac{|x|^n}{|x|^n + \kappa^n}. \tag{2.1.3}$$

The constant $\kappa > 0$ is termed the *half-maximal*, because $F(\kappa) = \frac{1}{2}$. The experiment will be considered thereafter with a parameter n typically greater than 1 (for instance $n = 1.7$).

The total current $I(t)$ corresponds to the integral of J in the interval $[0, L]$:

$$I(t) = I[\rho](t) = \int_0^L J(t,x)dx = J_0 \int_0^L \rho(x)F(c(t,x))dx. \tag{2.1.4}$$

Thus, the inverse problem consists in obtaining ρ from the measurement of $I[\rho](t)$, $\forall t \geq 0$.

We note that this is a Fredholm integral equation of the first kind, that is

$$I[\rho](t) = \int_0^L K(t,x)\rho(x)dx, \tag{2.1.5}$$

where $K(t,x) = J_0F(c(t,x))$ is called the kernel of the operator. This kind of problem is linear, but it is often ill-posed. For example, if K is sufficiently smooth, the operator defined above is compact on reasonable Banach spaces such as L^p spaces. Even if the mapping is injective, its inverse will not be continuous. Indeed, if I is compact and I^{-1} is continuous, then it follows that the identity map in L^p is compact, which is clearly false.

We study a simplified version of the above problem under more general assumptions that those considered in [38].

We assume the following

i)

$$c(t,x) = c_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right), \tag{2.1.6}$$

where erfc is the complementary error function:

$$\operatorname{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z \exp(-\tau^2)d\tau. \tag{2.1.7}$$

ii) We consider the following approximation of the Hill function (2.1.3)

$$F(x) \simeq F_m(x) = F(c_0)\sum_{j=1}^m a_j H(x-\alpha_j), \quad \forall x \in [0, c_0], \tag{2.1.8}$$

where H is the Heaviside unit step function i.e.

$$H(u) = \begin{cases} 1 & \text{if } u \geq 0, \\ 0 & \text{if } u < 0, \end{cases} \tag{2.1.9}$$

and a_j, α_j are positive constants such that:

$$\sum_{j=1}^m a_j = 1, \tag{2.1.10}$$

and

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m < c_0, \tag{2.1.11}$$

thus $\{\alpha_j\}_{j=1}^m$ defined a mesh on the interval $(0, c_0)$.

With the above assumptions we define the approximate total current

$$I_m[\rho](t) = J_0 \int_0^L \rho(x) K_m(t,x)dx, \tag{2.1.12}$$

where

$$K_m(t.x) = F_m(c(t,x)) = F_m\left(c_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)\right). \tag{2.1.13}$$

Therefore, our inverse problem is to recover ρ from the measurement $I_m[\rho](t)$, for all $t \geq 0$.

2.2 Main results

In this section we present the main results in this paper. We began to set the notation. For any $\gamma > 0$, we consider the function $\sigma_\gamma(x) = |x|^\gamma$, and introduce the following norms

$$\begin{aligned} \|f\|_{0,\gamma,b} &= \|\sigma_\gamma f\|_{L^2(0,b)}, \\ \|f\|_{1,\gamma,b} &= \|\sigma_\gamma f\|_{H^1(0,b)}, \\ \|f\|_{-1,\gamma,b} &= \|\sigma_\gamma f\|_{H^{-1}(0,b)}, \end{aligned}$$

and we set

$$L_k = L/\beta_k \quad \text{for } k = 1, ..., m, \quad \text{and } L_0 = 0, \tag{2.2.1}$$

where

$$\beta_j = \operatorname{erfc}^{-1}(\alpha_j/c_0) 2\sqrt{D} \ \text{ for } j = 1, ..., m. \tag{2.2.2}$$

First, we present a result about the continuity of the operator I_m.

Theorem 2.2.1. *Let $\rho : [0, L] \to \mathbb{R}$ be a function in $L^2(0, L)$. Then for $\gamma \geq \frac{3}{4}$ there exists a positive constant $C_1 > 0$ such that*

$$\|I_m[\rho]\|_{1,\gamma,L_m^2} \leq C_1 \|\rho\|_{L^2(0,L)}, \tag{2.2.3}$$

where C_1, only depend of $L, \alpha_1, \alpha_{m-1}, \alpha_m, a_m$ and γ.

Now we present our main results. First we define the constant

$$\gamma_0 = \max\left\{0, \frac{\ln(a_m)}{\ln\left(\frac{\beta_m}{\beta_{m-1}}\right)} - \frac{1}{2}\right\}. \tag{2.2.4}$$

Theorem 2.2.2. *Let $\rho : [0, L] \to \mathbb{R}$ be a function in $L^2(0, L)$. Then, for all $\gamma > \max\{\gamma_0, 3/4\}$, there exists a positive constant $C_2 > 0$ such that*

$$\|\rho\|_{-1,\gamma+1,L} \leq C_2 \|I_m[\rho]\|_{1,\frac{\gamma}{2}-\frac{1}{4},L_m^2}, \tag{2.2.5}$$

where C_2, only depends on $L, \alpha_1, \alpha_{m-1}, \alpha_m, a_m$ and γ.

This theorem corresponds to the observability inequality for the functional I_m. In particular, it establishes the identifiability for our inverse problem.

To obtain our result, we begin by studying the following functional

$$\Phi_m[\varphi](t) = \frac{1}{J_0 F(c_0)} I_m[\rho](t^2), \quad \forall t > 0. \tag{2.2.6}$$

where

$$\varphi(x) = \int_0^x \rho(z)dz. \tag{2.2.7}$$

We can note that the identifiability for Φ_m is equivalent to the identifiability for I_m.

Firstly we discuss some identifiability results for the operator Φ_m. For the following result, corresponding to a measurement on a small interval, we need to do a strong assumption about the regularity of the functions φ.

Theorem 2.2.3 (Identifiability for analytic functions). *If $\varphi : [0, L] \to \mathbb{R}$ is an analytic function such that*

$$\Phi_m[\varphi](t) = 0, \ \ \forall t \in (0, \delta), \tag{2.2.8}$$

for some $\delta > 0$, then $\varphi \equiv 0$ in $[0, L]$.

Remark 2.2.4. *In Theorem 2.2.3, the result is still true when $\Phi_m[\varphi]$ vanishes on any small open interval.*

The second identifiability result requires less regularity for φ, but it is necessary to consider more information.

Theorem 2.2.5. *Let $\varphi : [0, L] \to \mathbb{R}$ be a function such that*

$$\Phi_m[\varphi](t) = 0, \ \ \forall t \in [0, L_m]. \tag{2.2.9}$$

Then $\varphi \equiv 0$ in $[0, L]$.

Remark 2.2.6. *The proof of Theorem 2.2.5 is algebraic and give us ideas for reconstruction and numerical approximation.*

Theorems 2.2.3 and 2.2.5 give us, identifiability results for the operator I_m.

Furthermore, we present some continuity and stability result.

Theorem 2.2.7. *Let $\varphi \in H^1(0, L)$ be a function. Then there exists a constant $\tilde{C}_1 > 0$, such that*

$$\|\Phi_m[\varphi]\|_{H^1(0,L_m)} \leq \tilde{C}_1 \|\varphi\|_{H^1(0,L)}, \tag{2.2.10}$$

where $\tilde{C}_1$ only depend on $L, \alpha_1, \alpha_{m-1}, \alpha_m, a_m$.

Theorem 2.2.8. *Let $\varphi \in H^1(0,L)$ be a function. Then for $\gamma > \gamma_0$, there exists a constant $\tilde{C}_2 > 0$ such that*

$$\|\varphi\|_{0,\gamma,L} \leq \tilde{C}_2 \|\Phi_m[\varphi]\|_{H^1(0,L_m)}, \tag{2.2.11}$$

where the constant $\tilde{C}_2$ only depends on $L, \alpha_1, \alpha_{n-1}, \alpha_n, a_n$, and γ.

In the proof of Theorem 2.2.5 we can see the role played by the mesh $\{\alpha_j\}_{j=1}^m$, for $j = 1, ..., m-1$. Therefore we study the special case where we set

$$\alpha_j = c_0 \mathrm{erfc}\left(\frac{\beta_0 \beta^j}{2\sqrt{D}}\right), \quad j = 1, ..., m. \tag{2.2.12}$$

with $\beta \in (0,1)$ and $\beta_0 > 0$ constants. This simplification is not very restrictive because $\{\alpha_j\}_{j=1}^m$ defines a mesh in the interval $[0, c_0]$, therefore (2.2.12) only prescribes a non-regular mesh in which we want to have the approximation (2.1.8). In what follows, I_m and Φ_m are denoted by $\tilde{I}_m$ and $\tilde{\Phi}_m$, respectively, when α_j is given by (2.2.12).

On the other hand, for the reconstruction we first introduce the function

$$g(t) = \frac{\tilde{I}_m[\rho](t^2/\beta_0^2) - \tilde{I}_m[\rho](L_m^2)}{J_0 F(c_0)}, \quad \forall t \in [0, \beta_0 L_m). \tag{2.2.13}$$

We want to obtain a reconstruction algorithm and a numerical implementation to recover the function ρ from the measurement of $\tilde{I}_m[\rho]$. We begin by recovering $\tilde{\varphi} : [0, L] \to \mathbb{R}$ such that

$$\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) = g(t), \quad \forall t \in [0, \beta_0 L_m). \tag{2.2.14}$$

Next, we define $g_1, g_2, ..., g_m$ in the following recurrent form

$$g_1(x) = \begin{cases} \dfrac{1}{a_m} g\left(\dfrac{x}{\beta^m}\right), & \text{if } x \in [\beta L, L), \\ 0, & \text{otherwise}, \end{cases} \tag{2.2.15}$$

and

$$g_{k+1}(x) = \begin{cases} \dfrac{1}{a_m}\left(g\left(\dfrac{x}{\beta^m}\right) - \displaystyle\sum_{j=1}^{k} a_{m-k-1+j}\, g_j\left(\dfrac{\beta^j x}{\beta^{k+1}}\right)\right), & \text{if } x \in [\beta^{k+1}L, \beta^k L), \\ 0, & \text{otherwise}, \end{cases} \tag{2.2.16}$$

for $k = 1, .., m-1$. Furthermore for $k \geq m$, we define

$$g_{k+1}(x) = \begin{cases} \dfrac{1}{a_m}\left(g\left(\dfrac{x}{\beta^m}\right) - \displaystyle\sum_{j=1}^{m-1} a_j\, g_{j+k-m+1}\left(\dfrac{\beta^j x}{\beta^m}\right)\right), & \text{if } x \in [\beta^{k+1}L, \beta^k L), \\ 0, & \text{otherwise}. \end{cases} \tag{2.2.17}$$

With the above definitions we have the following reconstruction result.

Theorem 2.2.9. *Let ρ be a function in $\mathcal{C}^0([0,L])$, and let g be defined by (2.2.13) and $\{g_j\}_{j\geq 1}$ be given by definitions (2.2.15)-(2.2.17). Then the function $\tilde{\varphi}$ defined by*

$$\tilde{\varphi}(x) = \begin{cases} \sum_{j=1}^{+\infty} g_j(x), & x \in (0,L], \\ \\ g(0), & x = 0, \end{cases} \tag{2.2.18}$$

is well defined and satisfies the equation

$$\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) = g(t), \quad \forall t \in [0, \beta_0 L_m]. \tag{2.2.19}$$

Furthermore ρ satisfies

$$\int_0^x \rho(z)dz = \tilde{\varphi}(x) + \frac{\tilde{I}_m[\rho](L_m^2)}{J_0 F(c_0)}, \quad \forall x \in [0,L]. \tag{2.2.20}$$

Now we will present a stability result for $\tilde{\Phi}_m$ for a more general norm.

Let us consider, for $0 \leq a < b < \infty$, a norm $\|\cdot\|_{[a,b)}$ for functions $f : [a,b) \to \mathbb{R}$, which is finite whenever $f \in C^1([a,b])$, and such that for any $f \in C^1([0,\infty))$ we have

i) If $[a_1, b_1) \subset [a,b)$, then
$$\|f\|_{[a_1,b_1)} \leq \|f\|_{[a,b)}\,, \tag{2.2.21}$$

ii) For any $\lambda > 0$, there exists a positive constant $C(\lambda)$ such that
$$\|g_\lambda\|_{[\lambda a, \lambda b)} \leq C(\lambda)\,\|f\|_{[a,b)}\,, \tag{2.2.22}$$
where $g_\lambda(x) = f(x/\lambda)$, and $C(\cdot)$ is a increasing function.

For example, we note that the L^p norms satisfies (2.2.21)-(2.2.22) for any $1 \leq p \leq +\infty$, with
$$C(\lambda) = \begin{cases} \lambda^{\frac{1}{p}} & p \in [1, +\infty) \\ 1 & p = \infty. \end{cases}$$
But it is more interesting to consider the boundary variation norm defined by:

$$\|f\|_{BV} = \|f\|_{L^\infty(a,b)} + \sup_{a \leq x_1 < \cdots < x_k < b} \sum_{j=1}^{k} |f(x_k) - f(x_{k-1})|\,, \tag{2.2.23}$$

which also satisfies (2.2.21)-(2.2.22) with $C(\lambda) = 1$.

This kind of norms are more general, because permit consider functions with less regularity, for example step functions.

Theorem 2.2.10. *Let $\rho \in \mathcal{C}^0([0,L])$ be a function and considering a norm satisfying (2.2.21-2.2.22) conditions. Then we have*

$$\|\varphi(\cdot) - \varphi(L)\|_{[\beta^{k+1}L, \beta^k L)} \leq C(\beta_0)\frac{C(\beta_m)}{a_m^{k+1}} \left\|\tilde{\Phi}_m[\varphi](\cdot) - \tilde{\Phi}_m[\varphi](L_m)\right\|_{[\beta^{k+1}L_m, L_m)}. \tag{2.2.24}$$

2.3 Introduction to inverse problem

An inverse problem is a general framework that is used to convert observed measurements into information about a physical object or system that we are interested in. For example, if we have measurements of the Earth's gravity field, then we might ask the question: ´´given the data that we have available, what can we say about the density distribution of the Earth in that area?". The solution to this problem (i.e. the density distribution that best matches the data) is useful because it generally tells us something about a physical parameter that we cannot directly observe. Thus, inverse problems are one of the most important, and well-studied mathematical problems in science and mathematics. Inverse problems arise in many branches of science and mathematics, including computer vision, machine learning, statistics, statistical inference, geophysics, medical imaging (such as computed axial tomography and EEG/ERP), remote sensing, ocean acoustic tomography, nondestructive testing, astronomy, physics and many other fields.

The field of inverse problems was first discovered and introduced by Soviet-Armenian physicist, Viktor Ambartsumian. While still a student, Ambartsumian thoroughly studied the theory of atomic structure, the formation of energy levels, and the Schrödinger equation and its properties, and when he mastered the theory of eigenvalues of differential equations, he pointed out the apparent analogy between discrete energy levels and the eigenvalues of differential equations. He then asked: given a family of eigenvalues, is it possible to find the form of the equations whose eigenvalues they are? Essentially Ambartsumian was examining the inverse Sturm-Liouville problem, which dealt with determining the equations of a vibrating string. This paper was published in 1929 in the German physics journal Zeitschrift für Physik and remained in oblivion for a rather long time. Describing this situation after many decades, Ambartsumian said, "If an astronomer publishes an article with a mathematical content in a physics journal, then the most likely thing that will happen to it is oblivion".

Inverse problems appear in several fields, including medical imaging, image processing, mathematical finance, astronomy, geophysics, nondestructive material testing and sub-surface prospecting. Typical inverse problems arise from asking simple questions "backwards". For instance, the simple question might be "If we know precisely the structure of the inner organs of a patient, what kind of X-ray images would we get from her?". The same question backwards is "Given a set of X-ray images of a patient, what is the three- dimensional structure of her inner organs?". This is the inverse problem of Computerized Tomography, or CT imaging.

Usually the inverse problem is more difficult than the simple question that it reverses. For example, even though the Earth's gravitational field is governed by Newton's law of gravitation, the inverse problem of finding sub-surface structures from minor variations of the gravitational field on the surface is extremely hard. Successful solution of inverse problems requires specially designed algorithms that can tolerate errors in measured data.

Inverse problems research concentrates on the mathematical theory and practical interpretation of indirect measurements. The study of inverse problems is an active area of modern applied mathematics and one of the most interdisciplinary field of science.

2.3.1 Basic definitions

The inverse problem in abstract framework can be formulated as the problem where we want to define the inverse functional Λ^{-1} such that

$$\text{Data or measures} \xrightarrow{\Lambda^{-1}} \text{Model parameters.}$$

The inverse problem is considered the inverse to the forward problem which relates the model parameters to the data that we observe:

$$\text{Model parameters} \xrightarrow{\Lambda} \text{Data or measures.}$$

The transformation from data to model parameters (or vice versa) is a result of the interaction of a physical system with the object that we wish to infer properties about. In other words, the transformation is the physics that relates the physical quantity (i.e. the model parameters) to the observed data.

In the typical situation we have the well-defined the forward operator

$$\begin{array}{rccc} \Lambda : & \mathcal{P} & \longrightarrow & \mathcal{D}, \\ & p & \longrightarrow & \Lambda(p) = d, \end{array}$$

where $\mathcal{P}$ and $\mathcal{D}$ are the parameter and datum spaces respectively. The study of this kind problems is divided into several parts:

a) Injectivity of Λ (*identifiability problem*).

b) Continuity of Λ and its inverse Λ^{-1} if it exists (*stability problem*).

c) What is the range of Λ? (*characterization*).

d) Formula to recover p from $\Lambda(p)$ (*reconstruction problem*).

e) Numerical algorithm to find an approximation of p (*numerical reconstruction*).

Inverse problems are typically ill posed, as opposed to the well-posed problems more typical when modeling physical situations where the model parameters or material properties are known. Of the three conditions for a well-posed problem suggested by Jacques Hadamard (existence, uniqueness, stability of the solution or solutions) the condition of stability is most often violated.

2.3.2 Examples

One central example of a linear inverse problem is provided by a Fredholm first kind integral equation:

$$\Lambda[p](x) = d(x) = \int_a^b k(x, y)p(y)dy.$$

For sufficiently smooth k the operator defined above is compact on reasonable Banach spaces such as L^p spaces. Even if the mapping is injective its inverse will not be continuous. (However, by the bounded inverse theorem, if the mapping is bijective, then the inverse will be bounded (i.e. continuous)). Thus small errors in the data d are greatly amplified in the solution p. In this sense the inverse problem of inferring p from measured d is ill-posed. To obtain a numerical solution, the integral must be approximated using quadrature, and the data sampled at discrete points. The resulting system of linear equations will be ill-conditioned.

Another example of inverse problem is the well-known problem, about electrical impedance. This problem called Calderon's problem was introduced by A. P. Calderón in 1980 (see the seminal article *On an inverse boundary value problem* [14]). The Calderon's problem consists in recovering the electrical conductivity of the medium from its map over the boundary (for more reference see [92]).

Calderón asked the question of whether it is possible to determine the electrical conductivity of a body by making current and voltage measurements at the boundary. Put in mathematical terms, given a bounded and non-zero function $\gamma(x)$ (possibly complex-valued), which models the electrical conductivity of Ω. Then a potential $u \in H^1(\Omega)$ satisfies the Dirichlet problem

$$\begin{array}{rcll} \mathrm{div}(\gamma\nabla u) & = & 0 & \text{in } \Omega, \\ u & = & f & \text{on } \partial\Omega, \end{array} \tag{2.3.1}$$

where $f \in H^{\frac{1}{2}}(\partial\Omega)$ is a given boundary voltage potential. The Dirichlet-to-Neumann (DN) map is defined by

$$\Lambda[\gamma](f) = \gamma\frac{\partial u}{\partial\nu}|_{\partial\Omega}, \tag{2.3.2}$$

and if we assume that

$$\Lambda[\gamma_1](f) = \Lambda[\gamma_2](f), \quad \forall f \in H^{\frac{1}{2}}(\partial\Omega),$$

implies that γ_1 and γ_2 are equal? The answers have been given in many cases (see for instance [57, 90, 71, 6]).

Moreover, it is well-known that this problem can be reduced to studying the set of Cauchy data for the Schrödinger equation

$$-\Delta u + qu = 0, \tag{2.3.3}$$

with the potential q given by:

$$q = \frac{\Delta\sqrt{\gamma}}{\sqrt{\gamma}}. \tag{2.3.4}$$

In the case when the dimension is greater than 3 and all conductivities are in $C^2(\Omega)$, we have the following positive result.

Theorem 2.3.1. (Sylvester-Uhlmann 1987, see [90])

If $\Lambda_{\gamma_1} = \Lambda_{\gamma_2}$*, then* $\gamma_1 = \gamma_2$ *in* Ω*.*

We define the set of Cauchy data for a bounded potential q by:

$$\widehat{C}_q = \left\{ \left(u|_{\partial\Omega}, \frac{\partial u}{\partial \nu}|_{\partial\Omega} \right) \mid -\Delta u + qu = 0 \text{ on } \Omega,\ u \in H^1(\Omega) \right\} \subset H^{\frac{1}{2}}(\partial\Omega) \times H^{-\frac{1}{2}}(\partial\Omega),$$

the problem corresponds to assume $\hat{C}_{q_1} = \hat{C}_{q_2}$ then implies that $\gamma_1 = \gamma_2$, (see for instance [90, 71, 12]).

2.4 Mathematical modeling of signal transduction in olfactory cilia

Understanding the olfactory system is an important challenge in neuroscience today. The process of smelling, called olfaction, requires the reliable transmission of information about chemical identity and concentration to the brain via electrical signals. Olfactory stimuli begin when an odorant molecules bind to a receptor on the cytoplasmic membrane of cilium. Each olfactory receptor neuron (ORN) has several of these long, thin cilia that project into the nasal mucus. The cilium membrane has two primary sets of ion channels. The cyclic-nucleotide-gated (CNG) channels allow a depolarizing influx of Ca^{2+} and Na^{+} through the cilium membrane. As Ca^{2+} accumulates in the cilium, it activates Ca^{2+}-gated Cl^{-} channels which we call Cl(Ca) channels. An efflux of Cl^{-} through these channels causes a further depolarization. These are the two primary currents that are induced by signal transduction events within the cilia of olfactory receptor neurons. This signal then goes to the rest of the olfactory receptor neuron (ORN) and is routed to other neurons in the brain (see Figures 2.1 and 2.2).

The distribution of the channels should be crucial in determining the kinetics of the neuronal response. Experimental procedures developed by Steven Kleene and Rick Flannery in the College of Medicine (University of Cincinnati) have produced data from which the distributions of CNG channels can be inferred using mathematical and computational procedures developed by Donald French (see [37]). An unexpected result is that the distribution of the CNG channels is not uniform as has usually been assumed. For instance, Lindemann [63] produced a steady state model of cilia behavior assuming a uniform distribution of CNG channels. Dougherty et al [26] and Suzuki et al [89] also assumed a uniform CNG channel distribution in their detailed models.

Sensory transduction of odors occurs in olfactory receptor neurons (ORNs). Each olfactory receptor neuron projects several long, thin cilia into the nasal mucus, which extends from tip of the dendritic knob of each ORN ([78] and [82]). In frogs, each cilium is 25 to 200 μm long and includes two segments. The proximal segment occupies roughly the 20 percent of the cilium nearest the basal body [78]. The distal segment (the remaining 80 percent of the length) is slightly smaller in diameter. The cilia contain molecules required for transduction of an odor stimulus, including odorant receptor proteins, the G-proteins G_{olf}, Olfactory system

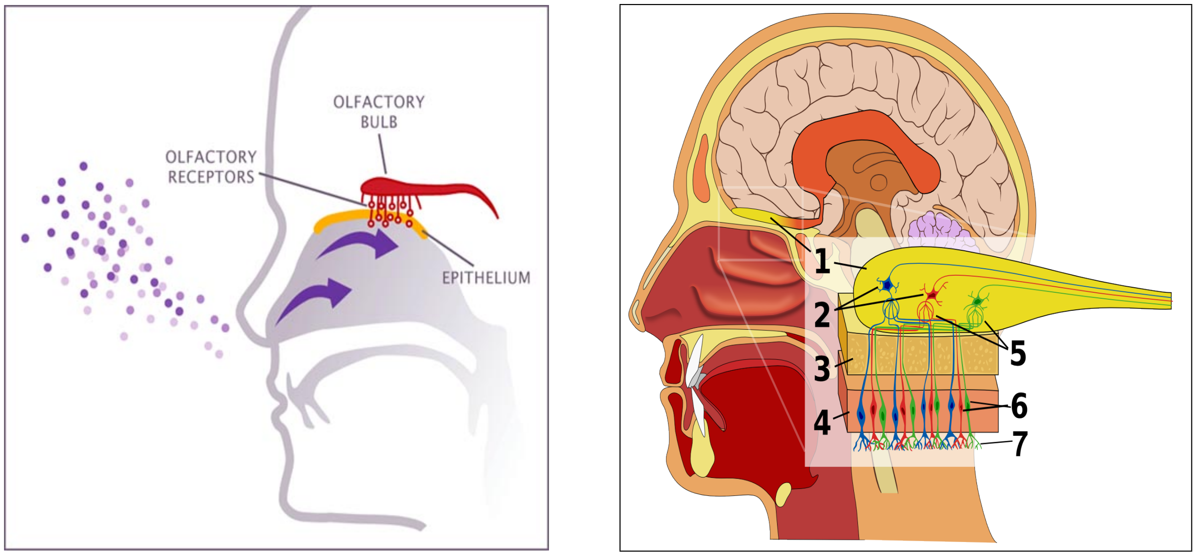


Figure 2.1: 1: Olfactory bulb, 2: Mitral cells, 3: Bone, 4: Nasal Epithelium, 5: Glomerulus, 6: Olfactory receptor cells and 7: Cilia in mucosa.

adenylyl cyclase III, intra-ciliary adenosine-3’, 5’, cyclic mono phosphate (cAMP) and two transduction channels ([82]).

Binding of an odor molecule to the membrane of a cilia activates the CNG and $Cl(Ca)$ currents as described in the introduction. The Ca^{2+} fraction of the inward flow through CNG channels is $> 80\%$ ([63]). Inward Cl^- current amplify the inward CNG current with little amplification of its current noise.

2.4.1 The experimental procedure

Experiments can be performed (see [18]) on single cilia; recordings are made of current as cAMP diffuses and activates the CNG channels. We explore the hypothesis that CNG channel distributions can be derived from the experimental current data and known properties of the cilia (a biological inverse problem). To accomplish this, we consider a mathematical model of the experiment. The model involves a system of nonlinear differential equations and a constrained Fredholm integral equation of the first kind which appears to be severely ill-conditioned.

The techniques for the procedure have been developed in [53, 55, 54, 56]. One olfactory cilium of a neuron is drawn into an open-ended recording pipette. The mouth of the pipette becomes attached to the base of the cilium. Then the pipette with the cilium inside is excised from the rest of the neuron. The base of the cilium remains attached to the open end of the pipette with the full length of the cilium inside the pipette. Both sides of the cilium are bathed in a Na^+-containing solution. cAMP is allowed to diffuse into the cilium from the bath surrounding the pipette. As the cAMP enters the cilium, some of the cAMP molecules bind to CNG channels. This allows the channels to open and initiate a transmembrane Na^+

A photograph of an olfactory receptor neuron

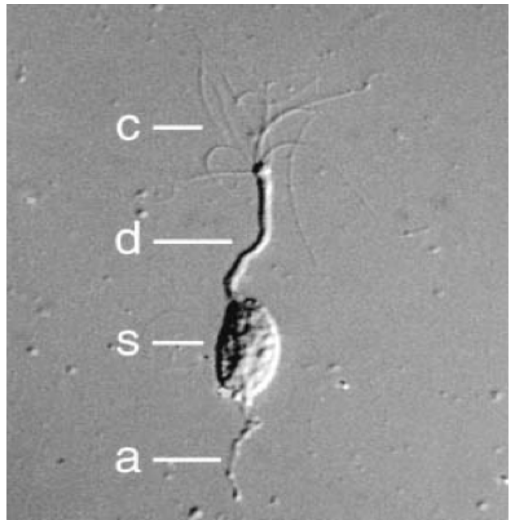


Figure 2.2: (C) Cilia, (D) Dendrite with a knob at the top, (S) Soma and (A) The axon.

current that is measured with electrodes placed inside and outside the pipette. The inside of the cilium is held at -50 mV, causing Na^+ ions to flow from the outside of the cilium to the inside. (Ca^{2+} is eliminated from all of the solutions and so is unavailable to gate the Ca^{2+}-activated Cl^- channels which are also known to exist in these cilia. Thus there is no Cl^- current.) See [18] for examples of experimental current versus time data.

The diffusion and binding of cAMP into the cilium is modeled by a nonlinear time-dependent partial differential equation which also depends on the channel distribution ρ. The membrane potential satisfies a second-order boundary value problem which depends on ρ and on the concentration of cAMP. The final equation in our model is the Fredholm integral equation which has a kernel that depends on the membrane potential and concentration of cAMP. It is constrained by the fact that the unknown (the channel distribution function ρ) must be nonnegative. Experimental work has shown that spatial variations in the concentration of Na^+ are insignificant.

2.4.2 Mathematical model

The following mathematical model was introduced by D. French, R. Flannery, C. Groetsch, W. Krantz and S. J. Kleene in their seminal article [37].

Frog olfactory cilia are thin tubes (diameter 2.8×10^{-5} cm [69] with lengths L in the range from 1.5×10^{-3} cm to as much as 20.0×10^{-3} cm) that are connected to the ends of the dendrites of the olfactory receptor neurons. We are mainly interested in the distribution of CNG channels $\rho = \rho(x)$. The units of ρ are number of CNG channels per unit length (in cm) and x is the distance from the base of the cilium where it is attached to the rest of the olfactory receptor neuron.

There are two key assertions about the binding of cAMP to the CNG channels in the presence of a local concentration of cAMP inside the cilium. These assumptions allow us to quantify the number of bound cAMP molecules and the current from the entrance of Na^+ ions. The radial dynamics are hypothesized to be insignificant; thus the problem is spatially one-dimensional in the length variable x, see Figure 2.3. The first assumption is

$$S(t,x) = B_s\rho(x)F(c(t,x)), \tag{2.4.1}$$

where $S(t,x)$ is the number of bound cAMP per unit length of the cilia on a thin ring at x and time t in the presence of cAMP concentration $c(t,x)$ and F is the classical dimensionless Hill's function, given by

$$F(x) = \frac{|x|^n}{|x|^n + \kappa^n}, \tag{2.4.2}$$

where the constant $\kappa > 0$ is the half-maximal, because $F(\kappa) = \frac{1}{2}$. In general $t \in [0,T]$ where T is a time in seconds when the experimental measurements are stopped. Typically the Hill's function parameters are the exponent $n = 1.7$ and half-maximal concentration $\kappa = 1.7$ μM. The function c is the concentration of cAMP in the thin ring in the cilium at (t,x) and has units of μM. The number B_s (its units are molecules/channel) measures the binding sites. Even though there are 4 cAMP binding sites on each CNG channel [68] the exponent n in the Hill's function suggests only $1-2$ sites are needed; we have chosen to take $B_s = n$.

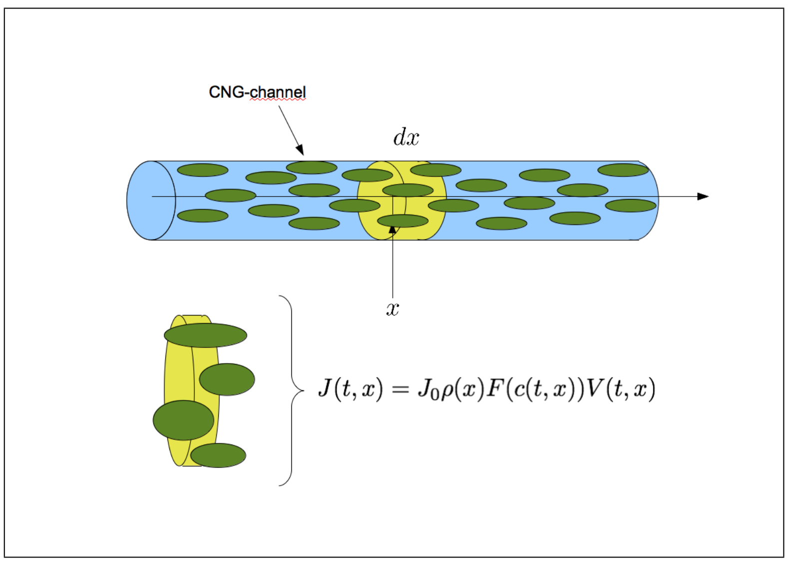


Figure 2.3: We approximated the cilium as a line segment.

The current flow assumption is

$$J(t,x) = J_0\rho(x)F(c(t,x))V(t,x), \tag{2.4.3}$$

(see [18]) where $J(t,x)$ is the Na^+ current through the CNG channels at point x on time t, and has units of pA/cm. The function $V(t,x)$ is the membrane potential, really part of the driving force with equilibrium potential equal to zero and has units of mV. The constant

$$J_0 = g_{CNG}\mathbb{P},$$

where the conductance is $g_{CNG} = 8.3 \times 10^{-12}$ S/channel and the constant $\mathbb{P} = .70$ is the maximum open probability for the channels.

Now we derive a key equation relating the total current at t, which we label $I(t)$, to $c(x,t)$ and $\rho(x)$. We assume that currents due to capacitive effects are negligible. Using the fact that

$$I(t) = \int_0^L J(t,x)dx = J_0 \int_0^L \rho(x)V(t,x)F(c(t,x))dx. \tag{2.4.4}$$

If ρ is the only unknown, this is a Fredholm integral equation of the first kind. These equations are well understood (see, e.g., [27, 47]), but the more complicated nonlinear system that arises in this way of modeling is non-standard. We must also impose the natural constraint that $\rho > 0$. Assuming that the rate of change of concentration of cAMP is balanced by Fick's law of diffusion of the cAMP and the binding, which is directly proportional to $\frac{\partial S}{\partial t}$ (see [25]), it follows, after matching units, that

$$\frac{\partial c}{\partial t} - D\frac{\partial^2 c}{\partial x^2} = -\alpha\frac{\partial S}{\partial t},$$

where the constant $\alpha = \frac{1}{N_A A}\ 2.7 \times 10^{-6} \frac{\mu\text{mol cm}}{\text{liter molecule}}$ and $D = 2.7 \times 10^{-6}\text{cm}^2/s$. Here, N_A is Avogadro's number and A is the cross-sectional area of the cilium. We have chosen D to be smaller than the diffusion constant for ATP, a similar molecule to cAMP, and bigger than the values predicted in [18]. Using $F' \geq 0$ and (2.4.1), we obtain

$$\begin{cases} \dfrac{\partial c}{\partial t} - \dfrac{D}{1+\alpha B_s F'(c)}\dfrac{\partial^2 c}{\partial x^2} &= 0, \quad t > 0,\ x \in (0,L) \\ c(0,x) &= 0, \qquad x \in (0,L) \\ c(t,0) &= c_0, \qquad t > 0 \\ \dfrac{\partial c}{\partial x}(t,L) &= 0, \qquad t > 0, \end{cases} \tag{2.4.5}$$

The final equation, for the membrane potential, is derived using standard cable theory (see, for instance, [52]). Current between the two electrodes passes through two significant resistances in series: the resistance of the ciliary membrane, and a longitudinal resistance along the length of the solution filling the cilium. This second resistance varies with distance along the length of the cilium, as does the transmembrane voltage. We find that $V = V(x,t)$ satisfies

$$\begin{cases} \dfrac{1}{r_a}\dfrac{\partial^2 V}{\partial x^2}(t,x) &= J(t,x), \quad t > 0,\ x \in (0,L) \\ V(t,0) &= -V_0, \qquad t > 0 \\ \dfrac{\partial V}{\partial x}(t,L) &= 0, \qquad t > 0, \end{cases} \tag{2.4.6}$$

where $r_a = 1.49 \times 10^{11}$ (S cm)$^{-1}$ is the intracellular resistance to longitudinal current of the saline solution in the cilium (here, r_a is formed by dividing the intracellular resistivity

$R_i = 91.7\,\Omega\,\mathrm{cm}$ by the cross-sectional area A). We have assumed the usual capacitance term is negligible as well as the background conductance.

We note that this problem can be transformed into a different form. By integrating the differential equation involving the membrane potential (2.4.6) in x from 0 to L we obtain, using the fact that the flux at the right boundary is zero and incorporating the integral equation (2.4.4),

$$-\frac{1}{r_a}\frac{\partial V}{\partial x}(t,0) = I(t). \tag{2.4.7}$$

Thus, we can view our problem as having unknowns c, V and ρ and consisting of the system (2.4.5,2.4.6) and the new condition above (2.4.7). In this reformulation we would view the new condition (2.4.7) as compensating for the "extra" unknown ρ.

French et al. [37] develops a mathematical model for this inverse problem and a numerical approximation of the solution. In [36], they found a solution in an approximate sense, that is, they consider a perturbation of the problem, considering the parameter

$$\varepsilon = \left(\frac{\kappa}{c_0}\right)^n$$

small enough and the approximation

$$F(x) \simeq 1 - \frac{\varepsilon}{x^n}, \quad \forall x \in (0,1).$$

In [38], under some approximations, the authors propose a linear inverse problem in which they find an explicit solution and propose a method to find a numerical solution to the inverse problem, the main approach corresponds to

$$c(t,x) \simeq w(t,x) = c_0\,\mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right), \tag{2.4.8}$$

where erfc is the complementary error function, defined by

$$\mathrm{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}}\int_0^z \exp(-\tau^2)d\tau. \tag{2.4.9}$$

V and S are constant, and

$$F(w(t,x)) \simeq H(\beta^2 t - x^2),$$

where $w(t,x)$ represents the concentration of cAMP, H is the Heaviside unit step function centered at the origin

$$H(s) = \begin{cases} 0 & , \quad s < 0 \\ 1/2 & , \quad s = 0 \\ 1 & , \quad s > 0 \end{cases}$$

and $\beta = 2\sqrt{D}\,\mathrm{erfc}^{-1}(\kappa/c_0)$.

2.5 Useful tools: Mellin's transform

In this section we recall well-know facts about the Mellin Transform (the reader is referred to [91] Chapter VIII, for details).

We first introduce the notation $<\alpha,\beta>$ for the open strip of complex numbers $s=\sigma+it$, $t\in\mathbb{R}$ such that $\alpha<\sigma<\beta$.

Definition 2.5.1 (Mellin transform). *Let f be locally Lebesgue integrable over $(0,+\infty)$. The Mellin transform of f is defined by*

$$\mathcal{M}[f](s)=\int_0^{+\infty} f(x)x^{s-1}dx,\quad \forall s\in<\alpha,\beta>,$$

where $<\alpha,\beta>$ is the largest open strip in which the integral converges (it is called the fundamental strip).

Lemma 2.5.1. *Let f be locally Lebesgue integrable over $(0,+\infty)$. Then we have the following properties*

1. *Let $s_0\in\mathbb{R}$ be a constant, then for all s such that $s+s_0\in<\alpha,\beta>$, we have*

$$\mathcal{M}[f(x)](s+s_0)=\mathcal{M}[x^{s_0}f(x)](s).$$

2. *We consider $g(x)=f(\beta x)$ for $\beta\in\mathbb{R}$ be a constant, then*

$$\mathcal{M}[g](s)=\beta^{-s}\mathcal{M}[f](s).$$

Definition 2.5.2 (Mellin transform as operator in L^2). *For functions in $L^2(0,+\infty)$ we define a linear operator $\tilde{\mathcal{M}}$ as*

$$\begin{aligned}\tilde{\mathcal{M}}:L^2(0,+\infty) &\longrightarrow L^2(-\infty,+\infty),\\ f &\longrightarrow \tilde{\mathcal{M}}[f](s):=\tfrac{1}{\sqrt{2\pi}}\mathcal{M}[f](\tfrac{1}{2}-is).\end{aligned}$$

Theorem 2.5.2 (The Mellin inversion theorem). *The operator $\tilde{\mathcal{M}}$ is invertible with inverse*

$$\begin{aligned}\tilde{\mathcal{M}}^{-1}:L^2(-\infty,+\infty) &\longrightarrow L^2(0,+\infty),\\ \varphi &\longrightarrow \tilde{\mathcal{M}}^{-1}[\varphi](x):=\tfrac{1}{\sqrt{2\pi}}\textstyle\int_{-\infty}^{+\infty}x^{-\frac{1}{2}-is}\varphi(s)ds.\end{aligned}$$

Furthermore, this operator is an isometry, that is,

$$\left\|\tilde{\mathcal{M}}[f]\right\|_{L^2(-\infty,\infty)}=\|f\|_{L^2(0,\infty)},\quad \forall f\in L^2(0,+\infty).$$

2.6 Results for the operator Φ_m

This section is devoted to the proof of Theorems 2.2.3,2.2.5,2.2.7 and 2.2.8. These are the principal results for the operator Φ_m. First we find a more easy expression for Φ_m, using (2.1.12) we have

$$\begin{aligned} I_m[\rho](t) &= J_0 \int_0^L \rho(x) K_m(t,x)dx && (2.6.1)\\ &= J_0 F(c_0) \sum_{j=1}^m a_j \int_0^L \rho(x) H(c(x,t)-\alpha_j)dx && (2.6.2)\\ &= J_0 F(c_0) \sum_{j=1}^m a_j \int\limits_{G_j(t)\cap(0,L)} \rho(x)dx, && (2.6.3) \end{aligned}$$

with $G_j(t) = \{x \in \mathbb{R} : \;\; c(x,t) \geq \alpha_j\}$, since "erfc" function is decreasing, we have

$$G_j(t) = \left[0, \beta_j \sqrt{t}\right], \tag{2.6.4}$$

where $\{\beta_j\}_{j=1}^m$ are given by (2.2.2) and $\beta_1 > \beta_2 > \cdots > \beta_m$. Thus we have

$$I_m[\rho](t) = J_0 F(c_0) \left(\sum_{j=1}^m a_j \int\limits_0^{h_j(\sqrt{t})} \rho(x)dx \right), \tag{2.6.5}$$

where $h_j(s) = \min\{L, \beta_j s\}$.

Replacing (2.6.5) in (2.2.6) and using (2.2.7) we obtain

$$\begin{aligned} \Phi_m[\varphi](t) &= \frac{1}{J_0 F(c_0)} I_m[\rho](t^2) \\ &= \sum_{j=1}^m a_j \varphi\left(h_j(t)\right) \quad \forall t \geq 0. \end{aligned} \tag{2.6.6}$$

In the rest of the paper we use (2.6.6) for the operator Φ_m.

Using the linearity of the operator Φ_m and since $\sum_{j=1}^m a_j = 1$, in the equation (2.6.39) we obtain

$$\Phi_m[f] = f, \;\; \text{for any constant function } f. \tag{2.6.7}$$

Proof of the Theorem 2.2.3.

Let φ be an analytic function such that

$$\Phi_m[\varphi](t) = \sum_{j=1}^n a_j \varphi(h_j(t)) = 0, \;\; \forall t \in (0,\delta),$$

then if we take $t \in (0, \min\{\delta, L_1\})$ and using that

$$L_0 < L_1 < \cdots < L_m, \tag{2.6.8}$$

we obtain $h_j(t) = \beta_j t,\ j = 1, ..., m$. Then we have

$$\Phi_m[\varphi](t) = 0 \Rightarrow \sum_{j=1}^{m} a_j \varphi(\beta_j t) = 0,$$

if we derive the above expression and evaluate at zero, we obtain

$$\varphi^{(k)}(0) \left(\sum_{j=1}^{m} a_j (\beta_j)^k \right) = 0, \ \ \forall k \geq 0,$$

where $\varphi^{(k)}(0)$ denotes the $k-$th derivative of φ at zero. But $\sum_{j=1}^{m} a_j (\beta_j)^k > 0$, since a_j, β_j are positive, we infer that $\varphi^{(k)}(0) = 0$, for all $k \geq 0$, then $\varphi \equiv 0$. This proves the identifiability for Φ_m in the case of analytic functions. □

To prove Theorem 2.2.5, we need some technical lemmas.

Lemma 2.6.1. *Let $f, g : [0, L] \to \mathbb{R}$ be functions, let $s, \alpha_0 \in [0, 1)$ and $\lambda \in (0, 1)$ be constants such that*

$$f(\tau) + g(\lambda \tau) = 0, \ \ \forall \tau \in [sL, L), \tag{2.6.9}$$

and

$$f(\tau) \equiv 0, \ \ \forall \tau \in [\alpha_0 L, L). \tag{2.6.10}$$

Then

$$g(\tau) = 0, \quad \forall \tau \in [\alpha_1 L, \lambda L), \tag{2.6.11}$$

where $\alpha_1 = \lambda \max\{s, \alpha_0\}$.

Proof. From (2.6.9-2.6.10), we obtain for $\tau \in [sL, L) \cap [\alpha_0 L, L) = [\max\{s, \alpha_0\} L, L)$

$$g(\lambda \tau) = 0.$$

This gives (2.6.11). □

Lemma 2.6.2. *Let $f : [0, L] \to \mathbb{R}$ be a function, let $s, \alpha_0 \in [0, 1)$ and $\lambda \in (0, 1)$ be real constants such that*

$$f(\tau) = 0, \ \ \forall \tau \in [\tilde{\alpha}_k L, L), \ \ \forall k \geq 1, \tag{2.6.12}$$

where

$$\tilde{\alpha}_k = \lambda \max\{s, \tilde{\alpha}_{k-1}\}, \ \ \forall k \geq 1, \tag{2.6.13}$$

with $\tilde{\alpha}_0 = \alpha_0$ given.

Then if $s > 0$,

$$f(\tau) = 0, \ \ \forall \tau \in [s\lambda L, L),$$

and if $s = 0$,

$$f(\tau) = 0, \ \ \forall \tau \in (0, L).$$

Proof. We need to consider two cases: $s = 0$ and $s > 0$.

If $s > 0$, actually $\exists k_0$, such that $\tilde{\alpha}_{k_0} < s$, and using (2.6.13) we obtain the result, because

$$\tilde{\alpha}_k = \lambda s, \quad \forall k > k_0.$$

Otherwise if $\tilde{\alpha}_k \geq s, \ \ \forall k$, replacing in (2.6.13), we have

$$\tilde{\alpha}_{k+1} = \lambda \tilde{\alpha}_k,$$

then $\tilde{\alpha}_k = \tilde{\alpha}_0 \lambda_1^k \to 0$, but that is impossible because $s > 0$.

Now if $s = 0$, replacing in (2.6.13) we obtain

$$\tilde{\alpha}_k = \alpha_0 \lambda^k.$$

Then using (2.6.12) we have

$$f(\tau) = 0, \ \forall \tau \in (0, L).$$

That completes the proof. □

Lemma 2.6.3. *Let $f : [0, L] \to \mathbb{R}$ be a function, let $s, \alpha_0 \in [0, 1)$, $\lambda_1, ...\lambda_n \in (0, 1)$ and $a_k > 0$, $k = 0, .., n$, such that $\lambda_1 > \lambda_2 > \cdots > \lambda_n \geq \alpha_0$, and*

$$a_0 f(t) + \sum_{j=1}^{n} a_j f(\lambda_j t) = 0, \ \ \forall t \in [sL, L), \tag{2.6.14}$$

and

$$f(\tau) \equiv 0, \ \ \forall \tau \in [\alpha_0 L, L). \tag{2.6.15}$$

Then

$$f(\tau) = 0, \quad \forall \tau \in [\overline{\alpha} L, L), \tag{2.6.16}$$

where $\overline{\alpha} = \lambda_n s$.

Proof. The proof is done by induction on n. In fact:

Case $n = 1$. In this case, from (2.6.14) we have the following equations

$$a_0 f(t) + a_1 f(\lambda_1 t) = 0 \ \ \forall t \in [sL, L), \tag{2.6.17}$$

$$f(\tau) = 0, \ \ \forall \tau \in [\alpha_0 L, L), \tag{2.6.18}$$

and $\alpha_0 \leq \lambda_1$. Then if we apply Lemma 2.6.1, with $g = f$, we obtain

$$f(\tau) = 0 \ \ \forall \tau \in [\alpha_1 L, \lambda_1 L),$$

where $\alpha_1 = \lambda_1 \max\{\alpha_0, s\}$, and we have

$$f(\tau) = 0 \ \ \forall \tau \in [\alpha_1 L, L),$$

because $\alpha_0 \leq \lambda_1$.

If $\alpha_0 = 0$, we obtain the desired result

$$f(\tau) = 0, \ \ \forall \tau \in [\lambda_1 sL, L),$$

on the other hand, when $\alpha_0 > 0$, we can apply again Lemma 2.6.1 but with α_1, because we have

$$a_0 f(t) + a_1 f(\lambda_1 t) = 0 \ \ \forall t \in [sL, L),$$
$$f(\tau) = 0, \ \ \forall \tau \in [\alpha_1 L, L),$$

and $\alpha_1 \leq \lambda_1$. Then we obtain

$$f(\tau) = 0, \ \ \forall \tau \in [\alpha_k L, L), \ \ \forall k \geq 1, \tag{2.6.19}$$

where

$$\alpha_k = \lambda_1 \max\{s, \alpha_{k-1}\}, \ \ \forall k \geq 1, \tag{2.6.20}$$

but if $s = 0$, replacing $t = 0$, in (2.6.17) we obtain $f(0) = 0$, then using Lemma 2.6.2 with (2.6.19-2.6.20) we have

$$f(\tau) = 0, \ \forall \tau \in [\lambda_1 sL, L).$$

That completes the case $n = 1$.

Case $n + 1$, Let us consider the induction hypothesis: For any $1 > \lambda_1 > \lambda_2 \cdots > \lambda_n$, $a_i > 0$, $i = 0, ..., n$, $\alpha_0 \leq \lambda_n$, are such that

$$a_0 f(t) + \sum_{j=1}^{n} a_j f(\lambda_j t) = 0, \ \ \forall t \in [sL, L),$$

and

$$f(\tau) \equiv 0, \ \ \forall \tau \in [\alpha_0 L, L).$$

Then

$$f(\tau) \equiv 0, \ \ \forall \tau \in [\lambda_n sL, L).$$

We want to prove that if $1 > \lambda_1 > \lambda_2 \cdots > \lambda_{n+1}$, $a_i > 0$, $i = 0, ..., n$, $\alpha_0 \leq \lambda_{n+1}$, are such that

$$a_0 f(t) + \sum_{j=1}^{n+1} a_j f(\lambda_j t) = 0, \ \ \forall t \in [sL, L), \tag{2.6.21}$$

and

$$f(\tau) \equiv 0, \ \ \forall \tau \in [\alpha_0 L, L). \tag{2.6.22}$$

Then

$$f(\tau) \equiv 0, \ \ \forall \tau \in [\lambda_{n+1} sL, L).$$

In fact, let us define the function

$$\psi(\tau) = \sum_{j=1}^{n+1} a_j f(\frac{\lambda_j}{\lambda_1}\tau) = a_1 f(\tau) + \sum_{j=2}^{n+1} a_j f(\tilde{\lambda}_j \tau),$$

where $\tilde{\lambda}_j = \frac{\lambda_j}{\lambda_1}$, $j = 2, .., n+1$, and we have $1 > \alpha_0/\lambda_{n+1}$, then using (2.6.22) we have

$$\psi(\tau) = 0, \ \ \forall \tau \in [\lambda_1 \frac{\alpha_0}{\lambda_{n+1}} L, L). \tag{2.6.23}$$

On the other hand, from (2.6.21), we have

$$a_0 f(\tau) + \psi(\lambda_1 \tau) = 0 \ \ \forall \tau \in [sL, L),$$

then from (2.6.22) and Lemma 2.6.1, with $g = \psi$, we conclude that

$$\psi(\tau) = 0, \ \ \forall \tau \in [\lambda_1 \max\{\alpha_0, s\} L, \lambda_1 L).$$

Now we set $s_1 = \lambda_1 \max\{\alpha_0, s\}$, and using (2.6.23) we have $\psi \equiv 0$, on $[s_1 L, \lambda_1 L) \cup [\lambda_1 \frac{\alpha_0}{\lambda_{n+1}} L, L)$, therefore

$$\psi(\tau) = a_1 f(\tau) + \sum_{i=2}^{n+1} a_i f(\tilde{\lambda}_i \tau) = 0, \ \ \forall \tau \in [s_1 L, L). \tag{2.6.24}$$

We can note that, $1 > \tilde{\lambda}_2 > \tilde{\lambda}_3 > \cdots > \tilde{\lambda}_{n+1}$, and $\alpha_0 \leq \lambda_{n+1} < \frac{\lambda_{n+1}}{\lambda_1} = \tilde{\lambda}_{n+1}$. Then if we use the induction hypothesis in (2.6.24) and (2.6.22), we obtain

$$f(\tau) = 0, \quad \forall \tau \in [\tilde{\alpha}_1 L, L),$$

where $\tilde{\alpha}_1 = s_1 \tilde{\lambda}_{n+1} = \lambda_{n+1} \max\{s, \alpha_0\} < \lambda_{n+1}$. Then we can repeat the last argument replacing α_0 by $\tilde{\alpha}_1$, and we have

$$f(\tau) = 0, \ \ \forall \tau \in [\tilde{\alpha}_k L, L), \ \ \forall k \geq 1,$$

where

$$\tilde{\alpha}_k = \lambda_{n+1} \max\{s, \tilde{\alpha}_{k-1}\}, \ \ \forall k \geq 1, \tag{2.6.25}$$

with $\tilde{\alpha}_0 = \alpha_0$ given. If $s = 0$, we replace $t = 0$, in (2.6.21) we have $f(0) = 0$, and using Lemma 2.6.2 we obtain

$$f(\tau) = 0, \ \ \forall \tau \in [\overline{\alpha} L, L),$$

where $\overline{\alpha} = \lambda_{n+1} s$, which completes the proof. $\square$

Proof of the Theorem 2.2.5.

Let $\varphi : [0, L] \to \mathbb{R}$ be a function such that

$$\Phi_m[\varphi](t) = \sum_{j=1}^{m} a_j \varphi(h_j(t)) = 0, \ \ \forall t \in [0, L_m],$$

then if $t = L_m$, we obtain

$$h_j(L_m) = L, \ \ \forall j = 1, ..., m,$$

thus

$$0 = \Phi_m[\varphi](L_m) = \varphi(L). \tag{2.6.26}$$

Then if $t \in [L_{k-1}, L_k]$, for $k = 1, .., m$, we obtain

$$\sum_{j=k}^{m} a_j \varphi(\beta_j t) = 0, \ \ \forall t \in [L_{k-1}, L_k],$$

that is equivalent to

$$a_k \varphi(t) + \sum_{j=k+1}^{m} a_j \varphi\left(\frac{\beta_j}{\beta_k} t\right) = 0, \ \ \forall t \in [\beta_k L_{k-1}, \beta_k L_k] = [\beta_k L_{k-1}, L], \tag{2.6.27}$$

for $k = 1, 2, ..., m$. We want to prove that

$$\varphi(\tau) = 0 \ \ \forall \tau \in [\beta_m L_{k-1}, L],$$

for $k = 1, 2, ..., m$, we will proceed by induction on $i = m - k \in \{0, ..., m-1\}$.

Case $i = 0$. Let $k = m$, replacing in (2.6.27), we have

$$a_m \varphi(t) = 0, \ \ \forall t \in [\beta_m L_{m-1}, L],$$

that implies

$$\varphi(\tau) = 0, \ \ \forall \tau \in [\beta_m L_{m-1}, L]. \tag{2.6.28}$$

Which completes the case $i = 0$.

Case $i = 1$. We set $k = m - 1$, replacing in (2.6.27), we have

$$a_{m-1} \varphi(t) + a_m \varphi\left(\frac{\beta_m}{\beta_{m-1}} t\right) = 0, \ \ \forall t \in [\beta_{m-1} L_{m-2}, L], \tag{2.6.29}$$

then we infer from Lemma 2.6.3 with $\lambda_1 = \frac{\beta_m}{\beta_{m-1}}$, $s = \frac{\beta_{m-1}}{\beta_{m-2}}$ and $\alpha_0 = \frac{\beta_m}{\beta_{m-1}}$, that

$$\varphi(\tau) = 0, \ \ \forall \tau \in [\beta_m L_{m-2}, L].$$

Case i. If we assume the case $i - 1$ holds true, i.e., if we have

$$\varphi(\tau) = 0, \ \ \forall \tau \in [\beta_m L_{m-i}, L], \tag{2.6.30}$$

thus replacing $k = m - i$ in (2.6.27), we have

$$a_{m-i} \varphi(t) + \sum_{j=m-i+1}^{m} a_j \varphi\left(\frac{\beta_j}{\beta_{m-i}} t\right) = 0, \ \ \forall t \in [\beta_{m-i} L_{m-i-1}, L], \tag{2.6.31}$$

then if we set $\lambda_j = \frac{\beta_j}{\beta_{m-i}} < 1$, for $j = n - i + 1, .., m$,

$$s = \beta_{m-i} \frac{L_{m-i-1}}{L}$$

and $\alpha_0 = \frac{\beta_m}{\beta_{m-i}} = \lambda_m$, then again Lemma 2.6.3, we obtain

$$\varphi(\tau) = 0, \ \ \forall \tau \in [\beta_m L_{m-i-1}, L].$$

Then

$$\varphi(\tau) = 0, \ \ \forall \tau \in [\beta_m L_{k-1}, L],$$

for $k = 1, .., m$, that implies

$$\varphi(\tau) = 0, \ \ \forall \tau \in [0, L].$$

This concludes the proof of Theorem 2.2.5. □

Now we will prove Theorems 2.2.7.

Proof of the Theorem 2.2.7.

We do some estimates,

$$\begin{aligned}
\|\varphi \circ h_j\|^2_{L^2(0,L_m)} &= \int_0^{L_m} \varphi^2(h_j(t))dt \\
&= \int_0^{L_j} \varphi^2(\beta_j t)dt + \varphi^2(L)L\left(\frac{1}{\beta_m} - \frac{1}{\beta_j}\right) \\
&\leq \frac{1}{\beta_j}\int_0^{L} \varphi^2(t)dt + \varphi^2(L)\frac{L}{\beta_m} \\
&\leq \frac{1}{\beta_m}\left\{\|\varphi\|^2_{L^2(0,L)} + \varphi^2(L)L\right\} \\
&\leq \frac{1}{\beta_m}\left(1 + (\mathrm{Tr}(L))^2 L\right)\|\varphi\|^2_{H^1(0,L)}, \qquad (2.6.32)
\end{aligned}$$

where

$$\mathrm{Tr}(u), \ \ \forall u > 0, \qquad (2.6.33)$$

is the norm of the trace operator in $H^1(0, u)$.

Now, if we set

$$c_1 = \frac{1}{\sqrt{\beta_m}}\left(1 + (\mathrm{Tr}(L))^2 L\right)^{\frac{1}{2}},$$

Then using (2.6.32) we obtain

$$\|\Phi_m[\varphi]\|_{L^2(0,L_m)} \leq \sum_{j=1}^{m} a_j \|\varphi \circ h_j\|_{L^2(0,L_m)} \leq c_1 \|\varphi\|_{H^1(0,L)}. \qquad (2.6.34)$$

Finally, let ψ be any test function with compact support in $(0, L_m)$. Then

$$\begin{aligned}\int_0^{L_m} \Phi_m[\varphi](t)\psi'(t)dt &= \sum_{j=1}^{m} a_j \left\{ \int_0^{L_j} \varphi(\beta_j t)\psi'(t)dt + \varphi(L) \int_{L_j}^{L_m} \psi'(t)dt \right\} \\ &= -\sum_{j=1}^{m} a_j\beta_j \int_0^{L_j} \varphi'(\beta_j t)\psi(t)dt \\ &= -\sum_{j=1}^{m} a_j\beta_j \int_0^{L_m} \varphi'(\beta_j t)\psi(t)(1 - H(\beta_j t - L))dt, \quad (2.6.35)\end{aligned}$$

thus

$$(\Phi_m[\varphi])'(t) = \sum_{j=1}^{m} a_j\beta_j\varphi'(\beta_j t)(1 - H(\beta_j t - L)), \quad \forall t \in (0, L_m). \quad (2.6.36)$$

Therefore, for any $\varphi \in H^1(0, L)$, the function $\Phi_m[\varphi]$ belongs to $H^1(0, L_m)$, with (2.6.36) we obtain

$$\left\|(\Phi_m[\varphi])'\right\|_{L^2(0,L_m)} \leq \sum_{j=1}^{m} a_j\sqrt{\beta_j}\left(\int_0^L (\varphi')^2(t)dt\right)^{1/2} \leq \sqrt{\beta_1}\left\|\varphi'\right\|_{L^2(0,L)}. \quad (2.6.37)$$

Combining (2.6.37) with equation (2.6.34), we obtain

$$\|\mathcal{I}[\varphi]\|_{1,0,L_m} \leq \tilde{C}_1 \|\varphi\|_{1,0,L},$$

where $\tilde{C}_1 = \sqrt{(c_1)^2 + \beta_1}$. The proof of Theorem 2.2.7 is complete. □

Next we will prove Theorem 2.2.8. To obtain this stability result, we need the following lemmas.

Lemma 2.6.4. *Let $\varphi : [0, +\infty] \to \mathbb{R}$ be a function such that supp$(\varphi) \subset [0, L)$. Then we have*

$$\Phi_m[\varphi](t) = \sum_{j=1}^{m} a_j\varphi(\beta_j t), \quad \forall t \geq 0, \quad (2.6.38)$$

where $\{\beta_j\}_{j=1}^m$ has been defined in (2.2.2).

Proof. Using (2.6.6) we have

$$\Phi_m[\varphi](t) = \sum_{j=1}^{m} a_j\varphi(h_j(t)), \ \forall t \geq 0,$$

where $h_j(t) = \min\{L, \beta_j t\}$. Then using supp$(\varphi) \subset [0, L)$, we have

$$\varphi(h_j(t)) = \varphi(\beta_j t),$$

this completes the proof. □

Lemma 2.6.5. *Let $\varphi : [0, L] \to \mathbb{R}$, and $f : [0, L_m] \to \mathbb{R}$ be functions such that*

$$\Phi_m[\varphi](t) = f(t), \quad \forall t \in [0, L_m]. \tag{2.6.39}$$

Then

$$\mathcal{M}[\tilde{f}](s) = \left(\sum_{j=1}^{m} a_j \beta_j^{-s}\right) \mathcal{M}[\tilde{\varphi}](s) \quad \text{for } s \text{ in the fundamental strip associated with } \tilde{\varphi}, \tag{2.6.40}$$

where

$$\tilde{f}(t) = \begin{cases} f(t) - f(L_m) & 0 \le t \le L_m, \\ \\ 0 & t \ge 0, \end{cases} \quad , \quad \text{and} \quad \tilde{\varphi}(t) = \begin{cases} \varphi(t) - \varphi(L) & 0 \le t \le L, \\ \\ 0 & t \ge L. \end{cases} \tag{2.6.41}$$

Proof. If we replace $t = L_m$ in (2.6.39), we have

$$\varphi(L) = f(L_m).$$

Using (2.6.7), we obtain

$$\Phi_m[\tilde{\varphi}](t) = \tilde{f}(t), \quad \forall t \ge 0. \tag{2.6.42}$$

On the other hand, using Lemma 2.6.4, we have

$$\Phi_m[\tilde{\varphi}](t) = \sum_{j=1}^{m} a_j \tilde{\varphi}(\beta_j t), \quad \forall t \ge 0,$$

and, for any $b > 0$, by Lemma 2.5.1, we have

$$\mathcal{M}[\tilde{\varphi}(bt)](s) = b^{-s} \mathcal{M}[\tilde{\varphi}](s),$$

thus

$$\mathcal{M}\left[\Phi_m[\tilde{\varphi}]\right](s) = \left(\sum_{j=1}^{m} a_j \beta_j^{-s}\right) \mathcal{M}[\tilde{\varphi}](s), \quad \forall s \in <\alpha, \beta> .$$

Using this and applying Mellin transform in (2.6.39) we obtain (2.6.40). ☐

Lemma 2.6.6. *Let $\varphi \in H^1(0, L)$ be a function. Then for any $\gamma > \gamma_0$, there exists a positive constant $C_\gamma > 0$, such that*

$$\|\varphi\|_{0,\gamma,L} \le C_\gamma \left\{ \|\Phi_m[\varphi]\|_{0,\gamma,L_n} + |\Phi_m[\varphi](L_m)| \right\}, \tag{2.6.43}$$

where C_γ only depends on $L, \beta_1, \beta_{m-1}, \beta_m, a_m$ and γ.

Proof. Let $\gamma > 0$ be a fixed constant. Using Lemma 2.6.5 and Lemma 2.5.1 we obtain

$$\left(\sum_{j=1}^{m} a_j \beta_j^{-(\gamma + \frac{1}{2} - is)}\right) \tilde{\mathcal{M}}[x^\gamma \tilde{\varphi}(x)](s) = \tilde{\mathcal{M}}\left[x^\gamma \Phi_m[\tilde{\varphi}](x)\right](s), \quad \forall s \in \mathbb{R}, \tag{2.6.44}$$

where $\tilde{\varphi}$ has been defined in (2.6.41). On the other hand

$$\begin{aligned}\left|\sum_{j=1}^{m} a_j \beta_j^{-(\gamma+\frac{1}{2}-is)}\right| &\geq a_m \beta_m^{-\gamma-\frac{1}{2}} - \left|\sum_{j=1}^{m-1} a_j \beta_j^{-(\gamma+\frac{1}{2}-is)}\right| \geq a_m \beta_m^{-\gamma-\frac{1}{2}} - \sum_{j=1}^{m-1} a_j \beta_j^{-(\gamma+\frac{1}{2})} \\ &\geq a_m \beta_m^{-\gamma-\frac{1}{2}} - \beta_{m-1}^{-(\gamma+\frac{1}{2})} = \beta_m^{-\gamma-\frac{1}{2}} \left(a_m - \left(\frac{\beta_{m-1}}{\beta_m}\right)^{-(\gamma+\frac{1}{2})}\right). \quad (2.6.45)\end{aligned}$$

Therefore, if we choose $\gamma > \gamma_0$, we have

$$\frac{\ln(a_m)}{\ln(\frac{\beta_m}{\beta_{m-1}})} < \gamma + \frac{1}{2},$$

then we can set

$$c_\gamma = \beta_m^{-\gamma-\frac{1}{2}} \left(a_m - \left(\frac{\beta_{m-1}}{\beta_m}\right)^{-(\gamma+\frac{1}{2})}\right) > 0,$$

and we obtain

$$\left|\sum_{j=1}^{m} a_j \beta_j^{-(\gamma+\frac{1}{2}-is)}\right| \geq c_\gamma > 0 \quad \forall s \in \mathbb{R}.$$

Thus, if we use the Mellin's isometry in (2.6.44), we have

$$c_\gamma \left\|\tilde{\varphi}\right\|_{0,\gamma,L} \leq \left\|\Phi_m[\tilde{\varphi}]\right\|_{0,\gamma,L_m}. \quad (2.6.46)$$

On the other hand,

$$\varphi(t) = \tilde{\varphi}(t) + \varphi(L), \quad \forall t \in (0, L),$$

$$\Phi_m[\varphi](t) = \Phi_m[\tilde{\varphi}](t) + \Phi_m[\varphi](L_m), \quad \forall t \in (0, L_m).$$

Thus, using the triangle inequality, we have

$$\|\varphi\|_{0,\gamma,L} \leq \|\tilde{\varphi}\|_{0,\gamma,L} + |\varphi(L)| \frac{L^{\gamma+1/2}}{\sqrt{2\gamma+1}},$$

and

$$\|\Phi_m[\tilde{\varphi}]\|_{0,\gamma,L_m} \leq \|\Phi_m[\varphi]\|_{0,\gamma,L_m} + |\Phi_m[\varphi](L_m)| \frac{L_m^{\gamma+1/2}}{\sqrt{2\gamma+1}}.$$

Replacing in (2.6.46) and observing that $\varphi(L) = \Phi_m[\varphi](L_m)$, we obtain

$$\begin{aligned}c_\gamma \left\|\varphi\right\|_{0,\gamma,L} &\leq \|\Phi_m[\tilde{\varphi}]\|_{0,\gamma,L_m} + c_\gamma |\varphi(L)| \frac{L^{\gamma+1/2}}{\sqrt{2\gamma+1}} \\ &\leq \|\Phi_m[\varphi]\|_{0,\gamma,L_m} + |\Phi_m[\varphi](L_m)| \frac{L_m^{\gamma+1/2}}{\sqrt{2\gamma+1}} + c_\gamma |\Phi_m[\varphi](L_m)| \frac{L^{\gamma+1/2}}{\sqrt{2\gamma+1}} \\ &= \|\Phi_m[\varphi]\|_{0,\gamma,L_m} + \left(\left(\frac{1}{\beta_m}\right)^{\gamma+1/2} + c_\gamma\right) \frac{L^{\gamma+1/2}}{\sqrt{2\gamma+1}} |\Phi_m[\varphi](L_m)|. \quad (2.6.47)\end{aligned}$$

Thus if we set
$$C_\gamma = \frac{1}{c_\gamma} \max\left\{ \left(\left(\frac{1}{\beta_n}\right)^{\gamma+1/2} + c_\gamma \right) \frac{L^{\gamma+1/2}}{\sqrt{2\gamma+1}}, 1 \right\},$$
we obtain (2.6.43). The proof of Lemma 2.6.6 is complete. □

Proof of the Theorem 2.2.8.

Using Lemma 2.6.6, we have that for any $\gamma > \gamma_0$, there exists a constant $C_\gamma > 0$, such that
$$\|\varphi\|_{0,\gamma,L} \leq C_\gamma \left\{ \|\Phi_m[\varphi]\|_{0,\gamma,L_m} + |\Phi_m[\varphi](L_m)| \right\} \leq C_\gamma \left((L_m)^\gamma + \mathrm{Tr}(L_m) \right) \|\Phi_m[\varphi]\|_{H^1(0,L_m)},$$
where $\mathrm{Tr}(L_m)$ has been defined in (2.6.33) . Thus if we set
$$\tilde{C}_2 = C_\gamma \left(\left(\frac{L}{\beta_n}\right)^\gamma + \mathrm{Tr}(L_m) \right),$$
we obtain (2.2.11). The proof of Theorem 2.2.8 is complete.

□

2.7 Proof of stability result

In this section we will proof Theorems 2.2.1 and 2.2.2.

Proof of the Theorem 2.2.1.

Let us fix any $\gamma > 0$ and $\rho : [0, L] \to \mathbb{R}$ be a function in $L^2(0, L)$. From (2.2.6) we obtain that
$$\begin{aligned} (x^\gamma I_m[\rho](x))' &= \gamma x^{\gamma-1} I_m[\rho](x) + x^\gamma (I_m[\rho](x))' \\ &= \gamma x^{\gamma-1} I_m[\rho](x) + \frac{x^{\gamma-\frac{1}{2}} J_0 F(c_0)}{2} (\Phi_m[\varphi])'(\sqrt{x}), \end{aligned} \tag{2.7.1}$$
where $\varphi(x) = \int\limits_0^x \rho(\tau) d\tau$.

Since
$$\int_0^{L_m^2} x^{2\gamma-1} \left((\Phi_m[\varphi])'(\sqrt{x}) \right)^2 dx = 2 \int_0^{L_m} \tau^{4\gamma-1} \left((\Phi_m[\varphi])'(\tau) \right)^2 d\tau = 2 \left\| (\Phi_m[\varphi])' \right\|^2_{0,2\gamma-\frac{1}{2},L_m},$$

we have

$$\begin{aligned}
\|I_m[\rho]\|^2_{1,\gamma,L_m^2} &\le \|I_m[\rho]\|^2_{0,\gamma,L_m^2} + \left(\gamma\|I_m[\rho]\|_{0,\gamma-1,L_m^2} + \frac{|J_0F(c_0)|}{\sqrt{2}}\|(\Phi_m[\varphi])'\|_{0,2\gamma-\frac{1}{2},L_m}\right)^2 \\
&\le \|I_m[\rho]\|^2_{0,\gamma,L_m^2} + 2\gamma^2\|I_m[\rho]\|^2_{0,\gamma-1,L_m^2} + (J_0F(c_0))^2\|(\Phi_m[\varphi])'\|^2_{0,2\gamma-\frac{1}{2},L_m} \\
&\le (L^2+2\gamma^2)\|I_m[\rho]\|^2_{0,\gamma-1,L_m^2} + (J_0F(c_0))^2\|(\Phi_m[\varphi])'\|^2_{0,2\gamma-\frac{1}{2},L_m}\,. \qquad (2.7.2)
\end{aligned}$$

On other hand, using (2.2.6) and the change of variable $\tau = x^2$, we have

$$\begin{aligned}
\|\Phi_m[\varphi]\|^2_{0,2\gamma-\frac{3}{2},L_m} &= \frac{1}{(F(c_0)J_0)^2}\int_0^{L_m} x^{4\gamma-3}\left(I_m[\rho](x^2)\right)^2 dx \\
&= \frac{1}{2\,(F(c_0)J_0)^2}\|I_m[\rho]\|^2_{0,\gamma-1,L_m^2}\,. \qquad (2.7.3)
\end{aligned}$$

Replacing (2.7.3) in (2.7.2), we obtain

$$\|I_m[\rho]\|^2_{1,\gamma,L_m^2} \le (L^2+2\gamma^2)2\,(F(c_0)J_0)^2\|\Phi_m[\varphi]\|^2_{0,2\gamma-\frac{3}{2},L_m} + (F(c_0)J_0)^2\|(\Phi_m[\varphi])'\|^2_{0,2\gamma-\frac{1}{2},L_m}\,,$$

and if we suppose that $\gamma \ge \frac{3}{4}$, with Theorem 2.2.7, we have

$$\|I_m[\rho]\|_{1,\gamma,L_m^2} \le \sqrt{2L^2+4\gamma^2}J_0F(c_0)L^{2\gamma-\frac{1}{2}}\|\Phi_m[\varphi]\|_{H^1(0,L_m)} \le \sqrt{2L^2+4\gamma^2}J_0F(c_0)L^{2\gamma-\frac{1}{2}}\tilde{C}_1\|\varphi\|_{H^1(0,L)}\,.$$

But we have from Cauchy-Schwarz inequality that $|\varphi(x)| \le \sqrt{L}\,\|\rho\|_{L^2(0,L)}$, and hence

$$\|\varphi\|^2_{H^1(0,L)} = \|\varphi\|^2_{L^2(0,L)} + \|\rho\|^2_{L^2(0,L)} \le (L^2+1)\|\rho\|^2_{L^2(0,L)}\,,$$

therefore for any $\gamma \ge \frac{3}{4}$, we have

$$\|I_m[\rho]\|_{1,\gamma,L_m^2} \le C_1\|\rho\|_{L^2(0,L)}\,,$$

where

$$C_1 = \sqrt{2L^2+4\gamma^2}J_0F(c_0)L^{2\gamma-\frac{1}{2}}\tilde{C}_1(L^2+1)^{1/2},$$

thus we obtain the first inequality in (2.2.3). The proof of Theorem 2.2.1 is complete. □

Proof of the Theorem 2.2.2.

Let ψ be any test function with compact support in $(0, L)$ and let γ be a positive constant. Set

$$g_\gamma(x) = x^\gamma\rho(x),$$

and

$$\varphi(t) = \int_0^t \rho(x)dx,$$

Derivating, it follows that

$$(x^{\gamma+1}\varphi(x))' = (\gamma+1)x^{\gamma}\varphi(x) + g_{\gamma+1}(x),$$

thus

$$\begin{aligned} < g_{\gamma+1}, \psi > &= \int_0^L g_{\gamma+1}(x)\psi(x)dx = \int_0^L \left((x^{\gamma+1}\varphi(x))' - (\gamma+1)x^{\gamma}\varphi(x)\right)\psi(x)dx \\ &= -\int_0^L \left(x^{\gamma+1}\varphi(x)\psi'(x) + (\gamma+1)x^{\gamma}\varphi(x)\psi(x)\right)dx, \end{aligned}$$

then we have

$$\begin{aligned} | < g_{\gamma+1}, \psi > | &\leq \left(\|\varphi\|_{0,\gamma+1,L} + (\gamma+1)\|\varphi\|_{0,\gamma,L}\right)\|\psi\|_{H^1(0,L)} \\ &\leq (L+\gamma+1)\|\varphi\|_{0,\gamma,L}\|\psi\|_{H^1(0,L)}. \end{aligned}$$

Therefore

$$\|g_{\gamma+1}\|_{H^{-1}(0,L)} \leq (L+\gamma+1)\|\varphi\|_{0,\gamma,L}. \tag{2.7.4}$$

Thus, using Lemma 2.6.6 we have that for any $\gamma > \max\{\gamma_0, \frac{3}{4}\}$, there exists a constant $C_\gamma > 0$, such that

$$\|\rho\|_{-1,\gamma+1,L} = \|g_{\gamma+1}\|_{H^{-1}(0,L)} \leq (L+\gamma+1)\,C_\gamma\left\{\|\Phi_m[\varphi]\|_{0,\gamma,L_m} + |\Phi_m[\varphi](L_m)|\right\}. \tag{2.7.5}$$

Using (2.2.6), we have

$$\Phi_m[\varphi](L_m) = \frac{1}{F(c_0)J_0} I_m[\rho](L_m^2).$$

Replacing in (2.7.5) and using (2.7.3) change $2\gamma - 3/2$ by γ, we obtain

$$\begin{aligned} \|\rho\|_{-1,\gamma+1,L} &\leq \frac{(L+\gamma+1)}{|J_0F(c_0)|}C_\gamma\left\{\frac{1}{\sqrt{2}}\|I_m[\rho]\|_{0,\frac{\gamma}{2}-\frac{1}{4},L_m^2} + |I_m[\rho](L_m^2)|\right\} \\ &= \frac{(L+\gamma+1)}{\sqrt{2}|J_0F(c_0)|}C_\gamma\left\{\|I_m[\rho]\|_{0,\frac{\gamma}{2}-\frac{1}{4},L_m^2} + \frac{\sqrt{2}}{L_m^{\gamma-\frac{1}{2}}}|(L_m^2)^{\frac{\gamma}{2}-\frac{1}{4}}I_m[\rho](L_m^2)|\right\} \\ &\leq \frac{(L+\gamma+1)}{\sqrt{2}|J_0F(c_0)|}C_\gamma\left\{1 + \sqrt{2}\frac{\mathrm{Tr}(L_m^2)}{L_m^{\gamma-\frac{1}{2}}}\right\}\|I_m[\rho]\|_{1,\frac{\gamma}{2}-\frac{1}{4},L_m^2}, \end{aligned}$$

where Tr has been defined in (2.6.33). Therefore if we set

$$C_2 = \frac{(L+\gamma+1)}{\sqrt{2}|J_0F(c_0)|}C_\gamma\left\{1 + \sqrt{2}\frac{\mathrm{Tr}(L_m^2)}{L_m^{\gamma-\frac{1}{2}}}\right\},$$

we obtain (2.2.5). The proof of Theorem 2.2.2 is complete. □

2.8 Case $\beta_j = \beta_0 \beta^j$

This section is devoted to the proof of Theorems 2.2.9 and 2.2.10.

Proof of the Theorem 2.2.9.

Let ρ be a function in $\mathcal{C}^0([0,L])$, and we consider $\{g_j\}_{j\geq 1}$ have been defined in (2.2.15-2.2.17).

First we note that

$$\operatorname{Supp}(g_{k+1}) \subseteq [\beta^{k+1}L, \beta^k L), \quad \forall k \geq 0. \tag{2.8.1}$$

Then we can define the sequence $\{\varphi_p\}_{p\in\mathbb{N}}$ as

$$\varphi_p(x) = \sum_{j=1}^{p} g_j(x), \quad \forall x \geq 0.$$

Using (2.8.1) we have that $\forall x \in \ \mathbb{R} \setminus (0,L)$,

$$\varphi_p(x) = 0, \quad \forall p \in \mathbb{N},$$

then

$$\lim_{p\to+\infty} \varphi_p(x) = 0, \quad \forall x \in \mathbb{R} \setminus (0,L).$$

On the other hand, we consider the ceiling function

$$\lceil x \rceil = \min\{k \in \mathbb{Z} \bigm| k \geq x\},$$

i.e., it is the smallest integer not less than x.

Now we define

$$k^*(x) = \left\lceil \frac{\ln(x/L)}{\ln(\beta)} \right\rceil, \quad \forall x \in (0,L). \tag{2.8.2}$$

We have the following relation

$$x \in [\beta^{k^*(x)}L, \beta^{k^*(x)-1}L), \quad \forall x \in (0,L),$$

therefore we obtain for $x \in (0,L)$

$$\varphi_p(x) = g_{k^*(x)}(x), \quad \forall p \geq k^*(x),$$

thus

$$\lim_{p\to+\infty} \varphi_p(x) = g_{k^*(x)}(x), \quad \forall x \in (0,L). \tag{2.8.3}$$

This prove that $\tilde{\varphi}$ defined in (2.2.18) is well defined.

On the other hand, replacing (2.2.12) in (2.2.2) we obtain

$$\beta_j = \beta_0 \beta^j, \quad j = 1, \dots, m.$$

Using (2.8.1) we have

$$\mathrm{supp}(\tilde{\varphi}) \subset [0, L),$$

then with Lemma 2.6.4 we have

$$\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) = \sum_{j=1}^{m} a_j \tilde{\varphi}(\beta^j t), \quad \forall t \geq 0. \tag{2.8.4}$$

Replacing $t = 0$ and $t = \beta_0 L_m$ in (2.2.13), (2.8.4) and using (2.2.18), (2.2.15) we obtain

$$\tilde{\Phi}_m[\tilde{\varphi}](0) = g(0), \quad \text{and} \quad \tilde{\Phi}_m[\tilde{\varphi}](L_m) = g(\beta_0 L_m) = 0.$$

Now, if we take $t \in (0, \beta_0 L_m) = (0, L/\beta^m)$, we have

$$\beta^j t \in (0, \beta^{j-m} L), \quad \text{for } j = \{1, 2, ..., m\}.$$

We need consider two cases when $t < L$ and $t \geq L$:

Case $t < L$. We have

$$\beta^j t \in (0, L), \quad \text{for } j = \{1, 2, ..., m\},$$

and

$$k^*(\beta^j t) = j + k^*(t),$$

thus replacing (2.8.3) in (2.8.4) we obtain

$$\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) = \sum_{j=1}^{m} a_j g_{j+k^*(t)}(\beta^j t),$$

therefore using (2.2.17) with $k + 1 = m + k^*(t)$ we obtain

$$\begin{aligned}
\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) &= \sum_{j=1}^{m-1} a_j g_{j+k^*(t)}(\beta^j t) + a_m g_{m+k^*(t)}(\beta^m t) \\
&= \sum_{j=1}^{m-1} a_j g_{j+k^*(t)}(\beta^j t) + \left(g\,(t) - \sum_{j=1}^{m-1} a_j g_{j+k^*(t)} \left(\beta^j t\right) \right) \\
&= g(t).
\end{aligned}$$

Case $t \geq L$. We set

$$k_*(x) = \left\lfloor \frac{\ln(x/L)}{\ln(1/\beta)} \right\rfloor, \quad \forall x \geq L,$$

where

$$\lfloor x \rfloor = \max\{k \in \mathbb{Z} \;\big|\; k \leq x\},$$

is the floor function, i.e., it is the largest integer not greater than x.

Thus we have

$$\beta^{k_*(x)+1} x < L \leq \beta^{k_*(x)} x, \quad \forall x \geq L, \tag{2.8.5}$$

and

$$k^*(\beta^{k_*(t)+1}t) = 1, \tag{2.8.6}$$

then, we have

$$\beta^j t \geq L, \qquad \forall j \leq k_*(t),$$

$$\beta^j t < L, \quad \forall j \geq k_*(t)+1.$$

Using (2.8.4) and (2.8.3) we obtain

$$\begin{aligned}
\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) &= \sum_{j=k_*(t)+1}^{m} a_j g_{k^*(\beta^j t)}\left(\beta^j t\right) \\
&= \sum_{j=1}^{m-k_*(t)} a_{j+k_*(t)} g_{k^*(\beta^{j+k_*(t)}t)}\left(\beta^{j+k_*(t)}t\right).
\end{aligned}$$

And using (2.8.6) we have

$$k^*(\beta^{j+k_*(t)}t) = j \;\; \forall j \geq 1,$$

Therefore using (2.2.15-2.2.16) with $k+1 = m - k_*(t)$ we obtain

$$\begin{aligned}
\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) &= \sum_{j=1}^{m-k_*(t)} a_{j+k_*(t)} g_j\left(\beta^{j+k_*(t)}t\right) \\
&= \sum_{j=1}^{m-k_*(t)-1} a_{j+k_*(t)} g_j\left(\beta^{j+k_*(t)}t\right) + a_m g_{m-k_*(t)}\left(\beta^m t\right) \\
&= \sum_{j=1}^{m-k_*(t)-1} a_{j+k_*(t)} g_j\left(\beta^{j+k_*(t)}t\right) + \left(g(t) - \sum_{j=1}^{m-k_*(t)-1} a_{k_*(t)+j} g_j\left(\beta^{j+k_*(t)}t\right)\right) \\
&= g(t).
\end{aligned}$$

We obtain (2.2.19).

Now we will prove (2.2.20). Replacing (2.2.13) in (2.2.19) and using (2.2.6) we obtain

$$\begin{aligned}
\tilde{\Phi}_m[\tilde{\varphi}](t/\beta_0) &= \frac{\tilde{I}_m[\rho](t^2/\beta_0^2) - \tilde{I}_m[\rho](L_m^2)}{J_0 F(c_0)} \\
&= \tilde{\Phi}_m[\varphi](t/\beta_0) - \frac{\tilde{I}_m[\rho](L_m^2)}{J_0 F(c_0)}, \quad \forall t \in [0, \beta_0 L_m]\,,
\end{aligned} \tag{2.8.7}$$

where φ has been defined in (2.2.7). Using (2.6.7) and Theorem 2.2.5 we obtain (2.2.20), i.e.

$$\tilde{\varphi}(x) = \varphi(x) - \frac{\tilde{I}_m[\rho](L_m^2)}{J_0 F(c_0)}, \quad \forall x \in [0, L]\,.$$

This complete the proof of Theorem 2.2.9. $\square$

Proof of the Theorem 2.2.10.

Let ρ be a function in $\mathcal{C}^0([0,L])$, and we consider $\{g_j\}_{j\geq 1}$ have been defined in (2.2.15-2.2.17).

Using (2.2.20) with $x = L$ we obtain

$$\tilde{\varphi}(x) = \varphi(x) - \varphi(L), \quad \forall x \in [0,L], \tag{2.8.8}$$

where φ and $\tilde{\varphi}$ have been defined in (2.2.7) and (2.2.18) respectively.

Now we consider that the norm satisfies (2.2.21-2.2.22) and using (2.8.2-2.8.4) we obtain

$$\|\varphi(\cdot) - \varphi(L)\|_{[\beta^{k+1}L,\beta^k L)} = \|g_{k+1}\|_{[\beta^{k+1}L,\beta^k L)}\,. \tag{2.8.9}$$

We will prove that for any $k \geq 0$, we have

$$\|g_{k+1}\|_{[\beta^{k+1}L,\beta^k L)} \leq \frac{C(\beta_m)}{a_m^{k+1}} \|g\|_{[\beta^{k+1}\beta_0 L_m,\beta_0 L_m)}\,. \tag{2.8.10}$$

For prove (2.8.10) we will apply induction on k.

Case $k = 0$**.** Using (2.2.15) and the property (2.2.21-2.2.22), we obtain

$$\|g_1\|_{[\beta L,L)} \leq \frac{C(\beta_m)}{a_m} \|g\|_{[\beta\beta_0 L_m,\beta_0 L_m)}\,.$$

Case $k+1 \leq m$. Now we suppose that for all $n = 1, ..., k$, we have

$$\|g_n\|_{[\beta^n L,\beta^{n-1} L)} \leq \frac{C(\beta_m)}{a_m^n} \|g\|_{[\beta^n\beta_0 L_m,\beta_0 L_m)}\,. \tag{2.8.11}$$

And now we will prove (2.8.10) for $k+1$. Using (2.2.16) and the property (2.2.21-2.2.22), we obtain

$$\|g_{k+1}\|_{[\beta^{k+1}L,\beta^k L)} \leq \frac{1}{a_m}\left(C(\beta_m)\|g\|_{[\beta^{k+1}\beta_0 L_m,\beta^k\beta_0 L_m)} + \sum_{j=1}^{k} a_{m-k-1+j} C\left(\frac{\beta^{k+1}}{\beta^j}\right)\|g_j\|_{[\beta^j L,\beta^{j-1}L)}\right).$$

Using the induction hypothesis (2.8.11), we have

$$\begin{aligned}
&\|g_{k+1}\|_{[\beta^{k+1}L,\beta^k L)} \leq \\
&\frac{1}{a_m}\left(C(\beta_m)\|g\|_{[\beta^{k+1}\beta_0 L_m,\beta^k\beta_0 L_m)} + \sum_{j=1}^{k} a_{m-k-1+j} C\left(\frac{\beta^{k+1}}{\beta^j}\right)\frac{C(\beta_m)}{a_m^j}\|g\|_{[\beta^j\beta_0 L_m,\beta_0 L_m)}\right) \\
&\leq \frac{C(\beta_m)}{a_m^{k+1}}\left(a_m^k\|g\|_{[\beta^{k+1}\beta_0 L_m,\beta^k\beta_0 L_m)} + \sum_{j=1}^{k} a_{m-k-1+j} C\left(\frac{\beta^{k+1}}{\beta^j}\right)\|g\|_{[\beta^j\beta_0 L_m,\beta_0 L_m)}\right) \\
&\leq \frac{C(\beta_m)}{a_m^{k+1}}\left(a_m^k + \sum_{j=1}^{k} a_{m-k-1+j} C\left(\frac{\beta^{k+1}}{\beta^j}\right)\right)\|g\|_{[\beta^{k+1}\beta_0 L_m,\beta_0 L_m)}\,.
\end{aligned}$$

We note that

$$C(u) \leq 1, \quad \forall u \in (0,1), \tag{2.8.12}$$

because $C(\cdot)$ is decreasing. Therefore

$$C\left(\frac{\beta^{k+1}}{\beta^j}\right) \leq 1, \quad \forall j \in \{1, ..., k\},$$

Thus we obtain

$$\|g_{k+1}\|_{(0,L)} \leq \frac{C(\beta_m)}{a_m^{k+1}} \|g\|_{(\beta^{k+1}\beta_0 L_m, \beta_0 L_m)}.$$

That prove (2.8.10) for all $k = \{0, ..., m-1.\}$

Case $k+1 > m$**.** We supposed (2.8.11) for all $n \leq k$. Replacing (2.2.17) and using the property (2.2.21-2.2.22) and induction hypothesis, we obtain

$$\begin{aligned}
&\|g_{k+1}\|_{[\beta^{k+1}L, \beta^k L)} \leq \\
&\frac{1}{a_m}\left(C(\beta_m)\|g\|_{[\beta^{k+1}\beta_0 L_m, \beta^k \beta_0 L_m)} + \sum_{j=1}^{m-1} a_j C\left(\frac{\beta^m}{\beta^j}\right)\frac{C(\beta_m)}{a_m^{j+k-m+1}}\|g\|_{[\beta^{j+k-m+1}\beta_0 L_m, \beta_0 L_m)}\right) \\
&= \frac{C(\beta_m)}{a_m^{k+1}}\left(a_m^k \|g\|_{[\beta^{k+1}\beta_0 L_m, \beta^k \beta_0 L_m)} + \sum_{j=1}^{m-1} a_j C\left(\frac{\beta^m}{\beta^j}\right)\|g\|_{[\beta^{j+k-m+1}\beta_0 L_m, \beta_0 L_m)}\right) \\
&\leq \frac{C(\beta_m)}{a_m^{k+1}}\left(a_m^k + \sum_{j=1}^{m-1} a_j C\left(\frac{\beta_m}{\beta^j}\right)\right)\|g\|_{[\beta^{k+1}\beta_0 L_m, \beta_0 L_m)},
\end{aligned}$$

and using (2.8.12) we obtain $C(\beta^m/\beta^j) \leq 1$, for all $j \in \{1, ..., m-1\}$. This complete the proof of (2.8.10).

On the other hand, using (2.2.19) and (2.2.22), we have

$$\|g\|_{[\beta^{k+1}\beta_0 L_m, \beta_0 L_m)} = C(\beta_0)\left\|\tilde{\Phi}_m[\tilde{\varphi}]\right\|_{[\beta^{k+1}L_m, L_m)},$$

replacing (2.8.8) in (2.8.10) we obtain

$$\|g_{k+1}\|_{[\beta^{k+1}L, \beta^k L)} \leq \frac{C(\beta_m)}{a_m^{k+1}} C(\beta_0)\left\|\tilde{\Phi}_m[\varphi](\cdot) - \tilde{\Phi}_m[\varphi](L_m)\right\|_{[\beta^{k+1}L_m, L_m)},$$

and replacing in (2.8.9) we have (2.2.24). This completes the proof of Theorem 2.2.10.

□

2.9 Numerical results

Now we will discuss the numerical implementation of the scheme developed in the proof of Theorem 2.2.9.

First we choose $\{\alpha\}_{j=1}^{m}$ as in the equation (2.2.12), and we define

$$F_j = \begin{cases} 0 & j = 0, \\ F\left(\dfrac{\alpha_j + \alpha_{j+1}}{2}\right) & \forall j = 1, ..., m-1, \\ 1 & j = m, \end{cases}$$

then we fix a_j as follows:

$$a_j = F_j - F_{j-1}, \quad \forall j = 1, ..., m, \tag{2.9.1}$$

using that F is increasing and $0 \leq F(x) < 1$, for all $x \geq 0$, we have $a_j \geq 0$, for all $j = 1, ..., m$, and

$$\sum_{j=1}^{m} a_j = 1.$$

We can see the approximation Hill's function in Figure 2.4.

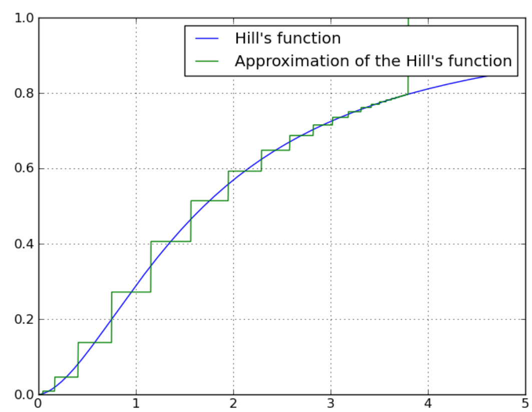


Figure 2.4: Hill's function and its approximation

We define a non-regular mesh in the interval $[0, L]$. We begin by defining

$$\mathcal{P}_{q,1} = \left\{ x_0 < x_1 < \cdots < x_q \middle| \; x_j \in [\beta L, L), \; x_0 = \beta L, \; x_{j-1} < x_j, \; \forall j = 1, ..., q \right\},$$

and its representative vector

$$\mathbf{P}_1 = \begin{pmatrix} x_0 \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_q \end{pmatrix} \in \mathbb{R}^{1+q}.$$

Thus

$$\mathcal{P}_{q,j} = \left\{ x \middle| \ \beta^{j-1} x \in \mathcal{P}_{q,1} \right\},$$

and their corresponding representative vector

$$\mathbf{P}_j = \beta^{j-1}\mathbf{P}_1 = \beta\mathbf{P}_{j-1} \in \mathbb{R}^{1+q}.$$

Thus if we fixed p, we want to recover in the mesh

$$\Sigma_{p,q} = \cup_{j=1}^{p} \mathcal{P}_{q,j},$$

and their corresponding representative vector is given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ . \\ . \\ . \\ \mathbf{P}_p \end{pmatrix} \in \mathbb{R}^{1+q+p}.$$

Using (2.2.15-2.2.17) we define the vectors $\mathbf{G}_1, \mathbf{G}_2, ..., \mathbf{G}_m \in \mathbb{R}^{q+1}$, in the following recurrent form

$$\left(\mathbf{G}_1\right)_s = g_1\left(\left(\mathbf{P}_1\right)_s\right), \quad \forall s = 1, ..., q+1, \tag{2.9.2}$$

and

$$\left(\mathbf{G}_{k+1}\right)_s = \frac{1}{a_m}\left(g\left(\frac{(\mathbf{P}_{k+1})_s}{\beta^m}\right) - \sum_{j=1}^{k} a_{m-k-1+j}(\mathbf{G}_j)_s \right), \quad \forall s = 1, ..., q+1, \tag{2.9.3}$$

for $k = 1, .., m-1$. Furthermore we define the vectors $\mathbf{G}_k \in \mathbb{R}^{1+q}$ for $k = m, ..., p-1$, as

$$(\mathbf{G}_{k+1})_s = \frac{1}{a_m}\left(g\left(\frac{(\mathbf{P}_{k+1})_s}{\beta^m}\right) - \sum_{j=1}^{m-1} a_j\left(\mathbf{G}_{j+k-m+1}\right)_s \right). \tag{2.9.4}$$

Thus we set the vector

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ . \\ . \\ . \\ \mathbf{G}_p \end{pmatrix} \in \mathbb{R}^{1+q+p},$$

this vector represent a discretization of the function $\tilde{\varphi}$ given by Theorem 2.2.9 on the mesh defined by $\Sigma_{p,q}$, i.e. $((\mathbf{P})_s, (\mathbf{G})_s)_{s=1}^{p+q+1}$ is a discretization of the curve $(x, \tilde{\varphi}(x))$, $x \in (0, L)$.

Therefore, using (2.2.20) and apply a forward difference schema we obtain an approximate of ρ, i.e. we defined the vectors $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{p+q}$

$$(\mathbf{X})_s = (\mathbf{P})_s, \quad (\mathbf{Y})_s = \max\left\{ \frac{(\mathbf{G})_{s+1} - (\mathbf{G})_s}{(\mathbf{P})_{s+1} - (\mathbf{P})_s}, 0 \right\}, \quad \forall s = 1, ..., p+q,$$

the maximum appears because we are looking for a density, i.e. a positive function. Thus $((\mathbf{X})_s, (\mathbf{Y})_s)_{s=1}^{p+q}$ is a approximation of the curve $(x, \rho(x))$, $x \in (0, L)$.

We consider the example considered in [38], thus we define

$$I(t) = \begin{cases} 0, & t \in (0, t_{Delay}), \\ I_{\text{Max}} \left[1 + \left(\dfrac{K_I}{t - t_{Delay}} \right)^{n_I} \right]^{-1}, & t > t_{Delay}, \end{cases} \tag{2.9.5}$$

with $t_{Delay} = 30ms$, $n_I \simeq 2.2$, $I_{Max} = 150pA$ and $K_I \simeq 100ms$. The current (2.9.5) is a sigmoidal function with short delay (Figure 2.5**B**), which is similar to the profiles in some applications (see [18], [33] or [58]).

A: approximation of ρ. **B**: current $I(t)$ as defined by (2.9.5)

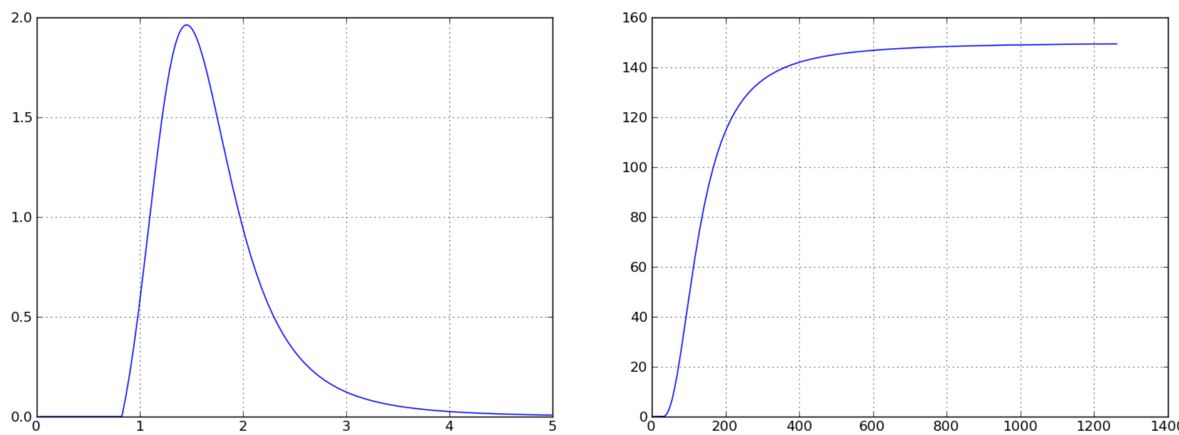


Figure 2.5: Approximation of the function $\rho(x)$ and with current $I(t)$ as defined by (2.9.5)

Thus we obtain the numerical solution as shown in Figure 2.5**A**. In this figure one sees that the main features of the solution of this model are consistent with those obtained in [33].

2.10 Polynomial approximation of the Hill's function

In this section we study a different approximation of the kernel (2.1.13), we will consider a polynomial approximation of the Hill's function (2.1.3), around c_0.

We consider P_m the Taylor's polynomial of degree m, from Hill's function defined by (2.1.3), at point $c_0 > 0$. Thus we have

$$F(x) = P_m(x - c_0) + O(|x - c_0|^{m+1}). \tag{2.10.1}$$

Therefore we define the polynomial kernel approximation by

$$PK_m(t.x) = P_m(w(t,x) - c_0), \tag{2.10.2}$$

where $w(t,x)$ is the solution of (2.1.1), given by

$$w(t,x) = c_0 - 2c_0\sqrt{2L}\left(\sum_{k=0}^{+\infty} \frac{e^{-\mu_k^2 Dt}}{\mu_k}\psi_k(x)\right), \tag{2.10.3}$$

with

$$\mu_k = \frac{2k+1}{2L}\pi, \tag{2.10.4}$$

and

$$\psi_k(x) = \left(\frac{2}{L}\right)^{1/2} \sin(\mu_k x). \tag{2.10.5}$$

Thus we define the total current with polynomial approximation, given by

$$PI_m[\rho](t) = \int_0^L \rho(x) PK_m(t,x)dx, \quad \forall t > 0. \tag{2.10.6}$$

Now we present our principal result released with the operator (2.10.6).

Theorem 2.10.1. *Let $m \leq 8$ be a constant. Then we have*

$$Ker\ PI_m = \{0\},$$

where

$$Ker\ PI_m = \left\{ f \in L^2(0,L) \middle|\ PI_m[f](t) = 0,\ \forall t > 0 \right\}.$$

Remark 2.10.2. *Theorem 2.10.1 corresponds to the identifiability for the operator PI_m.*

2.10.1 Proof of the identifiability Theorem 2.10.1

First we note that $\{\psi_k\}_{k\geq 0}$ is a orthonormal basis in $L^2(0,L)$. Thus for any $f \in L^2(0,L)$, we have

$$f(x) = \sum_{k\geq 0} < f, \psi_k > \psi_k(x), \quad \forall x \in (0,L), \tag{2.10.7}$$

where

$$< f, \psi_k >= \int_0^L f(x)\psi_k(x)dx.$$

If we set

$$G(t,x) = \sum_{k=0}^{+\infty} \frac{e^{-\mu_k^2 Dt}}{\mu_k}\psi_k(x), \tag{2.10.8}$$

we have

$$\begin{aligned}\|G(t,\cdot)\|^2_{L^2(0,L)} &= \sum_{k=0}^{+\infty}\frac{e^{-2\mu_k^2Dt}}{\mu_k^2}\|\psi_k\|^2_{L^2(0,L)}\\ &= \sum_{k=0}^{+\infty}\frac{e^{-2\mu_k^2Dt}}{\mu_k^2}\\ &\leq e^{-2\mu_0^2Dt}\left(\sum_{k=0}^{+\infty}\frac{1}{\mu_k^2}\right),\end{aligned}$$

using

$$\sum_{k=0}^{+\infty}\frac{1}{(2k+1)^2}=\frac{\pi^2}{8},$$

we obtain

$$\|G(t,\cdot)\|^2_{L^2(0,L)}\leq\frac{\pi^2}{8}e^{-2\mu_0^2Dt},\quad\forall t\geq 0. \tag{2.10.9}$$

On the other hand we have the following

Lemma 2.10.3. *Let* $\{a_k\}_{k\geq 0}$ *be a bounded sequence, i.e. exists a positive constant* $C>0$, *such that* $|a_k|\leq C$, *for all* $k\geq 0$. *Then we have that the series*

$$\sum_{k\geq 0}a_ke^{-\mu_k^2t}$$

is absolutely convergent for all $t>0$.

Proof. We need prove that the series is bounded, in fact

$$\begin{aligned}\sum_{k\geq 0}\left|a_ke^{-\mu_k^2t}\right| &\leq C\sum_{k\geq 0}\left(e^{-\frac{\pi^2t}{4L^2}}\right)^{(2k+1)^2}\\ &\leq C\sum_{k\geq 0}\left(e^{-\frac{\pi^2t}{4L^2}}\right)^{k}=\frac{C}{1-e^{-\frac{\pi^2t}{4L^2}}}.\end{aligned}$$

This completes the proof. □

Now, we define for all $t>0$,

$$\beta_k(t)=\frac{e^{-\mu_k^2Dt}}{\mu_k},\quad\forall k\geq 0,$$

and using Lemma 2.10.3 we have

$$\sum_{k\geq 0}\frac{1}{\mu_k}e^{-\mu_k^2Dt}=\sum_{k\geq 0}\beta_k(t),$$

is absolutely convergent for all $t>0$.

On other hand, we define the set

$$\Lambda_m = \left\{ \sum_{j=1}^{n} \mu_{n_j}^2 \,\Big|\, n_j \geq 0,\ \ n \leq m \right\}. \tag{2.10.10}$$

Let us set

$$P_m(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_m z^m,$$

where $\alpha_j \in \mathbb{R}$ for all $j = 0, 1, ..., m$. Let $\varepsilon > 0$ fixed constant, and for any $\rho \in L^2(0, L)$, we obtain, for all $s \geq 0$, that

$$\begin{aligned} PI_m[\rho](\varepsilon + s) &= \alpha_0 \int_0^L \rho(x)dx + \sum_{j=1}^{m} \alpha_j(-2c_0\sqrt{2L})^j \int_0^L \left(\sum_{k\geq 0} \frac{e^{-\mu_k^2 D(\varepsilon+s)}}{\mu_k} \Psi_k(x) \right)^j \rho(x)dx \\ &= \alpha_0 \int_0^L \rho(x)dx \\ &+ \sum_{j=1}^{m} \alpha_j(-2c_0\sqrt{2L})^j \sum_{k_1,\ldots,k_j \geq 0} e^{-\sum_{p=1}^{j} \mu_{k_p}^2 Ds} \int_0^L \Pi_{p=1}^{j} \beta_{k_p}(\varepsilon) \Psi_{k_p}(x)\rho(x)dx. \end{aligned}$$

Therefore we define $\{a_\lambda(\varepsilon, \rho)\}_{\lambda \in \Lambda_m}$, such that

$$PI_m[\rho](\varepsilon + s) = \alpha_0 \int_0^L \rho(x)dx + \sum_{\lambda \in \Lambda_m} a_\lambda(\varepsilon, \rho) e^{-\lambda D s},\ \ \forall s \geq 0. \tag{2.10.11}$$

Lemma 2.10.4. *Let $\rho \in L^2(0, L)$, be a function, such that*

$$PI_m[\rho](t) = 0,\ \ \forall t > 0. \tag{2.10.12}$$

Then we have

$$\int_0^L \rho(x)dx = 0, \tag{2.10.13}$$

and

$$a_\lambda(\varepsilon, \rho) = 0,\ \ \forall \lambda \in \Lambda_m. \tag{2.10.14}$$

Proof. First we see that the series defined in (2.10.11) is absolute convergent for all $s > 0$, in fact using Lemma 2.10.3, we only need to prove that $a_\lambda(\varepsilon, \rho)$ are bounded for all $\lambda \in \Lambda_m$, therefore we have

$$\begin{aligned} |a_\lambda(\varepsilon, \rho)| &\leq \sum_{j=1}^{m} |\alpha_j(-2c_0\sqrt{2L})^j \sum_{k_1,\ldots,k_j \geq 0} \left| \int_0^L \Pi_{p=1}^{j} \beta_{k_p}(\varepsilon) \Psi_{k_p}(x)\rho(x)dx \right| \\ &\leq \sum_{j=1}^{m} |\alpha_j(-2c_0\sqrt{2L})^j \left(\sum_{k\geq 0} |\beta_k(\varepsilon)| \right)^j \|\rho\|_{L^2(0,L)} < +\infty,\ \ \forall \lambda \in \Lambda_m, \end{aligned} \tag{2.10.15}$$

thus the series (2.10.11) is absolutely convergent for all $s > 0$.

We define $\{\lambda_k\}_{k\geq 1}$ as

$$\begin{aligned} \lambda_1 &= \min\Big\{\lambda \in \Lambda_m\Big\}, \\ \lambda_{k+1} &= \min\Big\{\lambda \in \Lambda_m \setminus \{\lambda_1, ..., \lambda_k\}\Big\}, \end{aligned} \tag{2.10.16}$$

we note that $0 < \lambda_1 < \lambda_2 < \cdots$.

Replacing in (2.10.11) we obtain

$$PI_m[\rho](2\varepsilon + s) = \alpha_0 \int_0^L \rho(x)dx + \sum_{k\geq 1} a_{\lambda_k}(\varepsilon,\rho)e^{-\lambda_k D\varepsilon}e^{-\lambda_k Ds}, \;\; \forall s \geq 0. \tag{2.10.17}$$

We set

$$S_j(s) = \sum_{k\geq j} a_{\lambda_k}(\varepsilon,\rho)e^{-\lambda_k D\varepsilon}e^{-\lambda_k Ds}, \;\; \forall s \geq 0,$$

and we have that exists a constant $C^{te} > 0$, such that for all $s > 0$, we have

$$|S_j(s)| \leq C^{te}e^{\lambda_j Ds}, \;\; \forall j \geq 1, \tag{2.10.18}$$

in fact,

$$|S_j(s)| \leq e^{-\lambda_1 Ds}\left(\sum_{k\geq 1} |a_{\lambda_k}(\varepsilon,\rho)|\, e^{-\lambda_k D\varepsilon}\right).$$

Replacing (2.10.17) in (2.10.12) we have

$$\alpha_0 \int_0^L \rho(x)dx + S_1(s) = 0, \;\; \forall s > 0, \tag{2.10.19}$$

thus take limit when s go to $+\infty$, and note that $\alpha_0 = F(c_0) \neq 0$, we obtain (2.10.13).

Now we want to prove (2.10.14), we apply induction on j.

Case n=1. We replacing (2.10.13) in (2.10.19) and multiplying by $e^{\lambda_1 Ds}$, we obtain

$$a_{\lambda_1}(\varepsilon,\rho)e^{-\lambda_1 D\varepsilon} + e^{\lambda_1 Ds}S_2(s) = 0, \;\; \forall s > 0, \tag{2.10.20}$$

but using (2.10.18) we have

$$|e^{\lambda_1 Ds}S_2(s)| \leq C^{te}e^{-(\lambda_2-\lambda_1)Ds},$$

thus note that $\lambda_1 < \lambda_2$ and take a limit when s go to 0, in (2.10.20) we obtain

$$a_{\lambda_1}(\varepsilon,\rho) = 0.$$

Case n+1. We assume that

$$a_{\lambda_j}(\varepsilon,\rho) = 0, \;\; \forall j = \{1, ..., n\}, \tag{2.10.21}$$

replacing (2.10.21) and (2.10.13) in (2.10.12) we obtain

$$a_{\lambda_{n+1}}(\varepsilon,\rho)e^{-\lambda_{n+1}D\varepsilon}+e^{\lambda_{n+1}Ds}S_{n+2}(s)=0,\ \ \forall s>0, \tag{2.10.22}$$

but using (2.10.18) we have

$$|e^{\lambda_{n+1}Ds}S_{n+2}(s)|\leq C^{te}e^{-(\lambda_{n+2}-\lambda_{n+1})Ds},$$

thus note that $\lambda_{n+1}<\lambda_{n+2}$ and take a limit when s go to 0, in (2.10.22) we obtain

$$a_{\lambda_{n+1}}(\varepsilon,\rho)=0.$$

Then we have (2.10.14).

This complete the proof of Lemma 2.10.4. □

Lemma 2.10.5. *If we consider* $\{\mu_k\}_{k\geq 0}$*, defined in* (2.10.4) *and we assumed that*

$$\mu_{n_1}^2+\cdots+\mu_{n_k}^2=\mu_n^2. \tag{2.10.23}$$

Then we have

$$k=1\ \ mod\ 8. \tag{2.10.24}$$

Proof. We have

$$\mu_n^2=\frac{\pi^2}{4L^2}(4\varphi(n)+1),$$

where $\varphi(n)=n^2+n$, and replacing in (2.10.23) we obtain

$$k+4\sum_{i=1}^{k}\varphi(n_i)=1+4\varphi(n),$$

using that $\varphi(p)$ is even for all $p\in\mathbb{N}$, we obtain (2.10.24). □

Proof of the Theorem 2.10.1.

Let $\rho\in L^2(0,L)$ be a function, from Lemma 2.10.4 we have (2.10.14) and if $m\leq 8$, using Lemma 2.10.5 we have that for all $k\geq 0$, doesn't exist $n_1,...,n_j$, such that

$$\mu_k^2=\sum_1^j\mu_{n_j}^2,$$

for any $j\leq m$, then for $\lambda=\mu_k^2\in\Lambda_m$, we obtain

$$a_{u_k^2}(\varepsilon,\rho)=e^{-\mu_k^2D\varepsilon}F'(c_0)(-2c_0\sqrt{2L})\int_0^L\psi_k(x)\rho(x)dx,$$

but $F'(c_0)\neq 0$, because F is increasing, then we have

$$<\psi_k,\rho>=0,\ \ \forall k\geq 0,$$

therefore from (2.10.7) we obtain

$$\rho=0.$$

This complete the proof of the Theorem 2.10.1.

□

2.11 Non-identifiability result in a non-linear inverse problem

We consider a variation of the olfactory inverse problem and we will show that, this inverse problem corresponds to determine a potential q in a interval of $\mathbb{R}$, from the Neumann data measured on one side of the boundary for the equation $-\Delta + q$. This is correlated with the problem of determining a complex-valued potential q in a bounded two dimensional domain, from the Cauchy data measured on an arbitrary open subset of the boundary for the associated Schrodinger equation $-\Delta + q$, the identifiability for this problem was shown in [48]. The motivation for this problem comes from a classical inverse problem of electrical impedance tomography. In this case one attempts to determine the electrical conductivity of a body, considering measurements of voltage and current on the boundary of the body. This problem was proposed by A. Calderón [14] and is well known as *Calderón's problem.* We proof the non-identifiability for our problem, this shows the importance of result [48]. Thus,

In order to simplify our problem, we will consider the domain as a cylinder, closed in the end and open in the other side, moreover, the boundary of de cylinder corresponds to the plasmatic membrane. Due to the symmetry of the domain, we assume a one dimensional model for the olfactory cilium, and we consider a diffusion-reaction model for the cAMP molecules, immersed in a solution. We denote by $w(t, x)$ the (volume) concentration of cAMP molecules. These molecules diffuse in the x-direction in a one dimensional channel of length L, thus we have the Neumman boundary condition $\frac{\partial c}{\partial x}(t, L) = 0$ for $t > 0$. Moreover, in our situation the concentration of cAMP is constant at the other side, that is, $c(t, 0) = c_0 > 0$ for $t > 0$, and initially the channel does not have cAMP.

Therefore, the conservation of mass is given by

$$\left\{ \begin{array}{rcll} \dfrac{\partial c}{\partial t} - D\dfrac{\partial^2 c}{\partial x^2} & = & -\alpha\dfrac{\partial S}{\partial t}, & t > 0,\ x \in (0, L) \\ c(0, x) & = & 0 & x \in (0, L) \\ c(t, 0) & = & c_0, & t > 0 \\ \dfrac{\partial c}{\partial x}(t, L) & = & 0 & t > 0, \end{array} \right. \tag{2.11.1}$$

where the constant $D > 0$, is the diffusion coefficient, $S(t, x)$ is the concentration the binding on the boundary for cAMP, and the constant $\alpha > 0$ is a unit conversion factor. Thus we have that

$$S = B_s\rho(x)F(c(t, x)), \tag{2.11.2}$$

where the constant $B_s > 0$ is the number of binding sites that are needed to activate the channel, $\rho(x)$ is the density of channels along the surface of the cilium and F is the classical dimensionless Hill function, given by

$$F(x) = \frac{|x|^n}{|x|^n + \kappa^n}, \tag{2.11.3}$$

where the constant $\kappa > 0$ is the half-maximal, because $F(\kappa) = \frac{1}{2}$. The constant $n > 0$, is a parameter, typically greater than 1.

The binding produces a change in the current flux $J(t, x)$ through the channels, which is governed by the following system:

$$\begin{aligned}
J(t,x) &= \mathbb{P}\frac{g_{CNG}}{B_S}S(t,x)v(t,x), & t>0,\ x\in(0,L) && (2.11.4)\\
\frac{1}{R_a}\frac{\partial^2 v}{\partial x^2} &= J(t,x), & t>0,\ x\in(0,L) && (2.11.5)\\
v(t,0) &= -v_0, & t>0 && (2.11.6)\\
\frac{\partial v}{\partial x}(t,L) &= 0, & t>0, && (2.11.7)
\end{aligned}$$

where the constant $\mathbb{P} > 0$, represents the probability of to find open the CNG channel, and the constants R_a and g_{CNG} are the resistance density of the longitudinal current in the cilium and the conductivity of CNG channel, respectively. We can see that the above system is parametric in the variable t, and the constant $v_0 > 0$ represents the voltage outside the cilium, furthermore the boundary condition at the end of the cilium is consistent with the fact that it is closed.

We can measure the quantity $I(t)$, which corresponds to the total current, which is simply the integral of J in the interval $[0, L]$:

$$I(t) = \int_0^L J(t,x)dx = -\frac{1}{R_a}\frac{\partial v}{\partial x}(0,t). \tag{2.11.8}$$

Our inverse problem consists in given the positive constants $L, D, w_0, B_s, n, \kappa, \mathbb{P}, g_{CNG}, R_a, v_0$, which define constitutive properties of the medium, if we measured $I(t)$ for all $t > 0$ (see 2.11.8), we are interested in to recover ρ, which is the density of channels along the surface of the cilium.

Let us define $q(x) = \mathbb{P}R_aF(w_0)\cdot\rho(x)$, and replace it in (2.11.5), then we obtain

$$v'' = qv.$$

Thus, we can reduce this inverse problem to following form: we want to determine the positive potential $q(x)$ such that:

$$\begin{cases} v'' - qv = 0, & x \in (0,L) \\ v(0) = -v_0, & \\ v'(L) = 0, & \end{cases} \tag{2.11.9}$$

and we measured $v'(0)$, where $q(x) = \mathbb{P}R_aF(w_0)\,\rho(x)$, and with this form, we obtain ρ.

Our first result corresponds to a non-identifiablity result.

Theorem 2.11.1. (Non-identifiability) *Given a constant* $v_0 > 0$*, there exist two different positives potential* $q_1, q_2 \in C^{1+\alpha}([0, L])$*, for some* $\alpha > 0$*, such that the respective solutions of 2.11.9* v_1, v_2*, satisfy*

$$v_1'(0) = v_2'(0). \tag{2.11.10}$$

Remark 2.11.2. *Apparently that is not exactly non-identifiability in the sense of the Calderón's problem, because we are trying to recover the potential, but only with one measure, and in the Calderon's problem, they try to recover the potential considering an arbitrary numbers of measures. If we call*

$$T_{q_j}(v_0) = v_j'(0), \quad j = 1, 2,$$

we note that T_{q_j} *is linear respect to* v_0*, then it follows from theorem 2.11.1, that there exist two different positive potentials,* $q_1, q_2 \in C^{1+\alpha}([0, L])$*, for some* $\alpha > 0$*, such that,*

$$T_{q_1}(v_0) = T_{q_2}(v_0), \quad \forall v_0,$$

it is exactly non-identifiability in the sense of the Calderón's problem in one dimension.

Remark 2.11.3. *We note that if we change the boundary condition at the end of the cilium, i.e., if we consider the following equation*

$$\begin{cases} v'' - qv &= 0, \quad x \in (0, L) \\ v(0) &= -v_0, \\ v(L) &= 0, \end{cases} \tag{2.11.11}$$

instead of (2.11.9)*, it is possible to obtain the same result as theorem 2.11.1*

We remark that this results holds in higher dimensions, considering that the domain admits separation of variables.

French et al, in [37], develop the mathematical model for this inverse problem and numerical approximation of solutions. In [36], they found a solution in an approximate sense, that is, they consider a perturbation of the problem, considering the parameter $\varepsilon = \left(\frac{\kappa}{w_0}\right)^n$ small enough and the approximation

$$F(x) \simeq 1 - \frac{\varepsilon}{x^n}, \quad \forall x \in (0, 1).$$

In [38], under some approximations, the authors propose a method to find a solution to the linearized inverse problem. The main difference in our analysis is that we consider the steady state for the cAMP, i.e., w is constant.

On the other hand, it is possible to obtain a positive identifiability result for the bidimensional case, in fact, let $\Omega \subseteq \mathbb{R}^2$ be a bounded domain with smooth boundary and let $\tilde{\Gamma}$ be a non-empty open subset of the boundary, we define $\Gamma_0 = \partial\Omega \setminus \tilde{\Gamma}$. Let ν be the unit outward normal vector to $\partial\Omega$, we denote $\frac{\partial u}{\partial \nu} = \nabla u \cdot \nu$, the normal derivative of u on $\partial\Omega$. Thus v is the solution of the following problem: 2.11.12

$$\begin{cases} \Delta v - qv &= 0, & \text{in } \Omega \\ v &= -v_0, & \text{on } \tilde{\Gamma} \\ v &= 0, & \text{on } \Gamma_0. \end{cases} \tag{2.11.12}$$

We can see that the boundary condition on Γ_0, corresponds to fix the potential at the end of the cilium. Hence, we have the following identifiability result.

Theorem 2.11.4. (Identifiability) *Let $q_1, q_2 \in C^{1+\alpha}(\overline{\Omega})$, with a positive constant $\alpha 0$, be two non-vanishing functions, which correspond to potential functions. Assume that for all $v_0 \in H^{\frac{1}{2}}(\partial\Omega)$ with supp$(v_0) \subset \tilde{\Gamma}$, the respective solutions of* (2.11.12), *v_1, v_2, satisfying*

$$\frac{\partial v_1}{\partial \nu} = \frac{\partial v_2}{\partial \nu} \quad \text{on } \tilde{\Gamma}. \tag{2.11.13}$$

Then $q_1 = q_2$.

We refer the readers to [48] for the proof of the above result.

2.11.1 About the Calderón's problem

In this section we will show some results related to the so-called Calderon's problem. In the seminal article *On an inverse boundary value problem* [14], A. P. Calderón asked the question of whether it is possible to determine the electrical conductivity of a body by making current and voltage measurements at the boundary. Put in mathematical terms, given a bounded and non-zero function $\gamma(x)$ (possibly complex-valued), which models the electrical conductivity of Ω. Then a potential $u \in H^1(\Omega)$ satisfies the Dirichlet problem

$$\begin{aligned} \operatorname{div}(\gamma \nabla u) &= 0 \quad \text{in } \Omega \\ u &= f \quad \text{on } \partial\Omega, \end{aligned} \tag{2.11.14}$$

where $f \in H^{\frac{1}{2}}(\partial\Omega)$ is a given boundary voltage potential. The Dirichlet-to-Neumann (DN) map is defined by

$$\Lambda_\gamma(f) = \gamma \frac{\partial u}{\partial \nu}|_{\partial\Omega}, \tag{2.11.15}$$

and if we assume that

$$\Lambda_{\gamma_1}(f) = \Lambda_{\gamma_2}(f) \;\; \forall f \in H^{\frac{1}{2}}(\partial\Omega)$$

implies that γ_1 and γ_2 are equal? The answers have been given in many cases (see for instance [57, 90, 71, 6]).

Moreover, it is well-known that this problem can be reduced to studying the set of Cauchy data for the Schrödinger equation

$$-\Delta u + qu = 0, \tag{2.11.16}$$

with the potential q given by:

$$q = \frac{\Delta\sqrt{\gamma}}{\sqrt{\gamma}}. \tag{2.11.17}$$

In the case when the dimension is greater than 3, Ω is a regular domain and all conductivities are in $C^2(\Omega)$, we have the following positive results.

Theorem 2.11.5. (Sylvester-Uhlmann 1987, see [90])

If $\Lambda_{\gamma_1} = \Lambda_{\gamma_2}$, then $\gamma_1 = \gamma_2$ in Ω.

We define the set of Cauchy data for a bounded potential q by:

$$\widehat{C}_q = \left\{ \left(u|_{\partial\Omega}, \frac{\partial u}{\partial\nu}|_{\partial\Omega} \right) \mid -\Delta u + qu = 0 \text{ on } \Omega,\ u \in H^1(\Omega) \right\} \subset H^{\frac{1}{2}}(\partial\Omega) \times H^{-\frac{1}{2}}(\partial\Omega),$$

the problem corresponds to assume $\hat{C}_{q_1} = \hat{C}_{q_2}$ then implies that $\gamma_1 = \gamma_2$, (see for instance [90, 71, 12]).

From the above section, we can see that our inverse problem is a type of the so-called Calderon's problem.

In [48], they show a result for global uniqueness from partial cauchy data in two dimensions. Their main result gives global uniqueness by measuring the Cauchy data on $\tilde{\Gamma} = \partial\Omega \setminus \overline{\Gamma_0}$. Let $q_j \in C^{1+\alpha}(\overline{\Omega})$, $j = 1, 2$ for some $\alpha > 0$ and let q_j be complex-valued. Consider the following sets of Cauchy data on an $\tilde{\Gamma}$:

$$C_{q_j} = \left\{ \left(u|_{\tilde{\Gamma}}, \frac{\partial u}{\partial\nu}|_{\tilde{\Gamma}} \right) \mid -\Delta u + q_j u = 0 \text{ on } \Omega,\ u|_{\Gamma_0} = 0,\ u \in H^1(\Omega) \right\}, \quad j = 1, 2.$$

Theorem 2.11.6. (Imanuvilov, Uhlmann & Yamamoto)

Let $\Omega \subset \mathbb{R}^2$ and assume $C_{q_1} = C_{q_2}$. Then $q_1 \equiv q_2$.

Using Theorem 2.11.6 one concludes immediately as a corollary the following global identifiability result for the conductivity equation 2.11.14.

Corollary 2.11.7. *With some $\alpha > 0$, let $\gamma_j \in C^{3+\alpha}(\overline{\Omega})$, $j = 1, 2$, be non-vanishing functions. Assume that*

$$\Lambda_{\gamma_1}(f) = \Lambda_{\gamma_2}(f) \;\; in\; \tilde{\Gamma} \;\; \forall f \in H^{\frac{1}{2}}(\partial\Omega), \;\; supp(f) \subset \tilde{\Gamma}.$$

Then $\gamma_1 \equiv \gamma_2$

2.11.2 Proof of the non identifiability result

In this section we will prove our first result.

Proof of Theorem 2.11.1.

We will prove the identifiability result. We note that the proof is for one dimensional case, but the proof in higher dimensions is similar to this one.

We define $m = v'(0)$, then multiplying (2.11.9) by v and integrating by parts, we have that,

$$v'(0)v_0 = \int_0^L \Big\{(v')^2 + q(x)v^2\Big\}dx \;\;\Rightarrow\;\; m \geq 0. \tag{2.11.18}$$

We will consider two cases:

1. If $m = 0$, then $v = -v_0$ for all $x \in [0, L]$, that implies that $q = 0$ for all $x \in [0, L]$.

2. If $m > 0$, we will looking for q constant, if $\lambda = \sqrt{q}$, then

$$v_1(x) = \frac{m}{\lambda}\sinh(\lambda x) - u_0\cosh(\lambda x) \tag{2.11.19}$$

is the solution of

$$\left\{\begin{array}{rcl} v'' - qv & = & 0 \\ v(0) & = & -v_0 \\ v'(0) & = & m. \end{array}\right. \tag{2.11.20}$$

with $q = \lambda^2$. We need to find $\lambda > 0$ such that $v_1'(L) = 0$ that is equivalent to

$$\tanh(\lambda L) = \frac{m}{\lambda u_0} \tag{2.11.21}$$

this equation has only one positive solutions $\lambda > 0$, then we choose $q_1 = \lambda^2$.

On the other hand, if $m > 0$, we will looking for q no constant, we will take $\tilde{L} = \min\{\frac{u_0}{m}, L\}$ and

$$v_2(x) = \begin{cases} -\frac{m}{2\tilde{L}}x^2 + mx - u_0 & x \leq \tilde{L} \\ \\ \frac{m\tilde{L}}{2} - u_0 & x > \tilde{L} \end{cases} \tag{2.11.22}$$

$$q_2(x) = \begin{cases} \frac{2\tilde{L}^2}{x^2 - 2\tilde{L}x + 2\frac{u_0\tilde{L}}{m}} & x \leq \tilde{L} \\ 0 & x > \tilde{L} \end{cases} \tag{2.11.23}$$

then v_2 is solution of (2.11.20) with $q = q_2$ and $v_2'(L) = 0$.

This completes the proof.

Chapter 3

Control of underwater vehicles in potential fluids

First we present a short preliminary background in essential topics. We consider a short introduction to Control Theory. This is essential to understand our results for our fluid-interaction control problem.

3.1 Control theory

We will focus on systems described in terms of ordinary differential equations. The control of partial differential equations remains out of our scope. The Control Theory is certainly, at present, one of the most interdisciplinary areas of research. Control Theory arises in most modern applications. The same could be said about the very first technological discoveries of the industrial revolution. On the other hand, Control Theory has been a discipline where many mathematical ideas and methods have melt to produce a new body of important Mathematics. Accordingly, it is nowadays a rich crossing point of Engineering and Mathematics. The word *control* has a double meaning. First, controlling a system can be understood simply as testing or checking that its behavior is satisfactory. In a deeper sense, to control is also to act, to put things in order to guarantee that the system behaves as desired. Let us indicate briefly how control problems are stated nowadays in mathematical terms. To fix ideas, assume we want to get a good behavior of a physical system governed by the state equation

$$A(y) = f(v) \tag{3.1.1}$$

Here, y is the state, the unknown of the system that we are willing to control. It belongs to a vector space Y . On the other hand, v is the control. It belongs to the set of admissible controls $\mathcal{U}_{ad}$. This is the variable that we can choose freely in $\mathcal{U}_{ad}$ to act on the system. Let us assume that $A : D(A) \subset Y \to Y$ and $f : \mathcal{U}_{ad} \to Y$ are two given (linear or nonlinear) mappings. The operator A determines the equation that must be satisfied by the state variable y, according to the laws of Physics. The function f indicates the way the control v acts on the system governing the state. For simplicity, let us assume that, for each $v \in \mathcal{U}_{ad}$, the state equation (3.1.1) possesses exactly one solution $y = y(v)$ in Y . Then, roughly speaking, to control (3.1.1) is to find $v \in \mathcal{U}_{ad}$ such that the solution to (3.1.1) gets close to the desired prescribed state. The "best" among all the existing controls achieving the desired goal is frequently referred to as the *optimal control*. This mathematical formulation might seem sophisticated or even obscure for readers not familiar with this topic. However, it is by now standard and it has been originated naturally along the history of this rich discipline. One of the main advantages of such a general setting is that many problems of very different nature may fit in it. As many other fields of human activities, the discipline of Control existed much earlier than it was given that name. Indeed, in the world of living species, organisms are endowed with sophisticated mechanisms that regulate the various tasks they develop. This is done to guarantee that the essential variables are kept in optimal regimes to keep the species alive allowing them to grow, develop and reproduce. Thus, although the mathematical formulation of control problems is intrinsically complex, the key ideas in Control Theory can be found in Nature, in the evolution and behavior of living beings. See [30] for a discussion on this and other related aspects.

This section focuses on the controllability of finite-dimensional control systems. That short introduction in control theory is based in the Jean-Michel Coron's Book, Control and Nonlinearity [24].

3.1.1 Controllability of linear finite dimensional systems

We will now be concerned with the controllability of ordinary differential equations. We will start by considering linear systems. We denote by $\mathcal{L}(\mathbb{R}^k;\mathbb{R}^l)$ the set of linear maps from $\mathbb{R}^k$ into $\mathbb{R}^l$. We often identify, in the usual way, $\mathcal{L}(\mathbb{R}^k;\mathbb{R}^l)$ with the set, denoted $\mathcal{M}_{k,l}(\mathbb{R})$, of $k \times l$ matrices with real coefficients. We denote by $\mathcal{M}_{k,l}(\mathbb{C})$ the set of $k \times l$ matrices with complex coefficients. Throughout this chapter, T_0, T_1 denote two real numbers such that $T_0 < T_1$, $A : (T_0, T_1) \to \mathcal{L}(\mathbb{R}^n;\mathbb{R}^n)$ denotes an element of $L^\infty((T_0,T_1);\mathcal{L}(\mathbb{R}^n;\mathbb{R}^n))$ and $B : (T_0,T_1) \to \mathcal{L}(\mathbb{R}^m;\mathbb{R}^n)$ denotes an element of $L^\infty((T_0,T_1);\mathcal{L}(\mathbb{R}^m;\mathbb{R}^n))$. We consider the time-varying linear control system

$$\frac{dx}{dt} = A(t)x + B(t)u, \quad t \in [T_0, T_1], \tag{3.1.2}$$

where, at time $t \in [T_0, T_1]$, the state is $x(t) \in \mathbb{R}^n$ and the control is $u(t) \in \mathbb{R}^m$.

Definition 3.1.1. *Let $b \in L^1((T_0,T_1);\mathbb{R}^n)$. A map $x : [T_0, T_1] \to \mathbb{R}^n$ is a solution of*

$$\frac{dx}{dt} = A(t)x + b(t), \quad t \in (T_0, T_1), \tag{3.1.3}$$

if $x \in \mathcal{C}^0([T_0,T_1];\mathbb{R}^n)$ and satisfies

$$x(t_2) = x(t_1) + \int_{t_1}^{t_2} (A(t)x(t) + b(t))\, dt, \quad \forall (t_1, t_2) \in [T_0, T_1]^2.$$

In particular, for $x^0 \in \mathbb{R}^n$, a solution to the Cauchy problem

$$\frac{dx}{dt} = A(t)x + b(t), \quad t \in (T_0, T_1), \ x(T_0) = x^0, \tag{3.1.4}$$

is a function $x \in \mathcal{C}^0([T_0, T_1]; \mathbb{R}^n)$ such that

$$x(\tau) = x^0 + \int_\tau^{T_0} (A(t)x(t) + b(t))\,dt, \quad \forall \tau \in [T_0, T_1]^2.$$

It is well known that, for every $b \in L^1((T_0, T_1); \mathbb{R}^n)$ and for every $x^0 \in \mathbb{R}^n$, the Cauchy problem (3.1.4) has a unique solution.

Let us now define the controllability of system (3.1.2).

Definition 3.1.2. *The linear time-varying control system* (3.1.2) *is controllable if, for every* $(x^0, x^1) \in \mathbb{R}^n \times \mathbb{R}^n$*, there exists* $u \in L^\infty((T_0, T_1); \mathbb{R}^m)$ *such that the solution* $x \in \mathcal{C}^0([T_0, T_1]; \mathbb{R}^n)$ *of the Cauchy problem* (3.1.4) *satisfies* $x(T_1) = x^1$*.*

Remark 3.1.1. *One could replace in this definition* $u \in L^\infty((T_0, T_1); \mathbb{R}^m)$ *by* $u \in L^2((T_0, T_1); \mathbb{R}^m)$ *or by* $u \in L^1((T_0, T_1); \mathbb{R}^m)$*. Spaces do not lead to different controllable systems. We have chosen to consider* $u \in L^\infty((T_0, T_1); \mathbb{R}^m)$ *since this is the natural space for general nonlinear control system.*

Let us first recall the definition of the resolvent of the time-varying linear system $x' = A(t)x$.

Definition 3.1.3. *The resolvent R of the time-varying linear system* $x' = A(t)x$ *is the map*

$$\begin{array}{rccc} R: & [T_0, T_1]^2 & \rightarrow & \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n) \\ & (t1, t2) & \mapsto & R(t1, t2), \end{array}$$

such that, for every $t_2 \in [T_0, T_1]$*, the map*

$$\begin{array}{rccc} R(\cdot, t_2): & [T_0, T_1] & \rightarrow & \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n) \\ & t1 & \mapsto & R(t1, t2), \end{array}$$

is the solution of the Cauchy problem

$$\dot{M} = A(t)M, \ M(t_2) = Id_n \tag{3.1.5}$$

where Id_n *denotes the identity map of* $\mathbb{R}^n$*.*

One has the following classical properties of the resolvent.

Proposition 3.1.2. *The resolvent R is such that*

$$R \in \mathcal{C}^0([T_0, T_1]^2; \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n)),$$

$$R(t_1, t_1) = Id_n, \ \forall t_1 \in [T_0, T_1],$$

$$R(t_1, t_2)R(t_2, t_3) = R(t_1, t_3), \ \forall (t_{1,t}2, t_3) \in [T_0, T_1]^3.$$

In particular,

$$R(t_1, t_2)R(t_2, t_1) = Id_n, \ \forall (t_1, t_2) \in [T_0, T_1]^2.$$

Moreover, if $A \in \mathcal{C}^0([T_0, T_1]; \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n))$*, then* $R \in \mathcal{C}^1([T_0, T_1]2; \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n))$ *and one has*

$$\frac{\partial R}{\partial t_1}(t, \tau) = A(t)R(t, \tau), \ \forall (t, \tau) \in [T_0, T_1]^2,$$

$$\frac{\partial R}{\partial t_2}(t, \tau) = -R(t, \tau)A(\tau), \ \forall (t, \tau) \in [T_0, T_1]^2.$$

Of course, the main property of the resolvent is the fact that it gives the solution of the Cauchy problem (3.1.4). Indeed, one has the following classical proposition.

Proposition 3.1.3. *The solution of the Cauchy problem* (3.1.4) *satisfies*

$$x(t_1) = R(t_1, t_0)x(t_0) + \int_{t_0}^{t_1} R(t_1, \tau)b(\tau)d\tau, \ \ \forall (t_0, t_1) \in [T_0, T_1]^2. \tag{3.1.6}$$

Equality (3.1.6) is known as Duhamel's principle.

In particular, the solution of (3.1.4) is given by

$$x(t) = R(t, T_0)x_0 + \int_{T_0}^{t} R(t, \tau)b(\tau)d\tau, \ \ \forall t \in [T_0, T_1].$$

An integral criterion for controllability: The controllability Gramian

We are now going to give a necessary and sufficient condition for the controllability of system (3.1.2) in terms of the resolvent of the time-varying linear system $x' = A(t)x$. This condition is due to Rudolph Kalman, Yu-Chi Ho and Kumpati Narendra [50].

Let us now define the controllability Gramian of the control system (3.1.2).

Definition 3.1.4. *The controllability Gramian of the control system*

$$\frac{dx}{dt} = A(t)x + B(t)u,$$

is the symmetric $n \times n$-matrix

$$\mathfrak{C} = \int_{T_0}^{T_1} R(T_1, \tau)B(\tau)B^t(\tau)R^t(T_1, \tau)d\tau. \tag{3.1.7}$$

In (3.1.7) and throughout this section, for a matrix M or a linear map M from $\mathbb{R}^k$ into $\mathbb{R}^l$, M^t denotes the transpose of M.

The controllability of (3.1.2) is given in the following theorem [50] due to Rudolph Kalman, Yu-Chi Ho and Kumpati Narendra.

Theorem 3.1.4. *The linear time varying control system* (3.1.2) *is controllable if and only if its controllability Gramian is invertible.*

The control given by the controllability Gramian $\mathfrak{C}$ is

$$\overline{u}(\tau) = B(\tau)^t R(T_1, \tau)^t \mathfrak{C}^{-1}(x^1 - R(T_1, T_0)x^0), \ \ \tau \in (T_0, T_1). \tag{3.1.8}$$

(In (3.1.8) and in the following, the notation "$\tau \in (T_0, T_1)$" stands for " for almost every $\tau \in (T_0, T_1)$ " or in the distribution sense in $D'(T_0, T_1)$, depending on the context.) Thus, Let $\overline{x} \in \mathcal{C}^0([T_0, T_1]; \mathbb{R}^n)$ be the solution of the Cauchy problem

$$\frac{d\overline{x}}{dt} = A(t)\overline{x} + b(t), \quad t \in (T_0, T_1), \ \ \overline{x}(T_0) = x^0. \tag{3.1.9}$$

Then, by Proposition 3.1.3, we have

$$\overline{x}(T_1) = x^1.$$

Proposition 3.1.5. *Let* $(x^0, x^1) \in \mathbb{R}^n \times \mathbb{R}^n$ *and let* $u \in L^2((T_0, T_1); \mathbb{R}^m)$ *be such that the solution of the Cauchy problem* (3.1.4) *satisfies*

$$x(T_1) = x^1.$$

Then if we consider $\overline{u}$ *defined in* (3.1.8)*, we have*

$$\int_{T_0}^{T_1} |\overline{u}(t)|^2 dt \leq \int_{T_0}^{T_1} |u(t)|^2 dt,$$

with equality (if and) only if

$$u(t) = \overline{u}(t) \quad \text{for almost every } t \in (T_0, T_1).$$

Kalman's type conditions for controllability

The necessary and sufficient condition for controllability given in Theorem 3.1.4 requires computing the matrix $\mathfrak{C}$, which might be quite difficult (and even impossible) in many cases, even for simple linear control systems. In this section we give a new criterion for controllability which is much simpler to check. For simplicity, we start with the case of a time invariant system, that is, the case where $A(t)$ and $B(t)$ do not depend on time. Note that a priori the controllability of $x' = Ax + Bu$ could depend on T_0 and T_1 (more precisely, it could depend on $T_1 - T_0$). This is indeed the case for some important linear partial differential control systems;

So we shall speak about the controllability of the time invariant linear control system $x' = Ax + Bu$ on $[T_0, T_1]$.

The famous Kalman rank condition for controllability is given in the following theorem.

Theorem 3.1.6. *The time invariant linear control system* $x' = Ax + Bu$ *is controllable on* $[T_0, T_1]$ *if and only if*

$$Span\left\{A^i Bu \ \ : \ u \in \mathbb{R}^m, \ i \in \{0, ..., n-1\}\right\} = \mathbb{R}^n.$$

In particular, whatever $T_0 < T_1$ and $T_0' < T_1'$ are, the time invariant linear control system $x' = Ax + Bu$ is controllable on $[T_0, T_1]$ if and only if it is controllable on $[T_0', T_1']$.

The authors of [50] say that Theorem 3.1.6 is the "simplest and best known" criterion for controllability.

Let us now turn to the case of time-varying linear control systems. We assume that A and B are of class $\mathcal{C}^\infty$ on $[T_0, T_1]$. Let us define, by induction on i a sequence of maps $B_i \in \mathcal{C}^\infty([T_0, T_1]; \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n))$ in the following way:

$$B_0(t) := B(t), \ \ B_i(t) := \dot{B}_{i-1}(t) - A(t)B_{i-1}(t), \ \ \forall t \in [T_0, T_1]. \tag{3.1.10}$$

Then one has the following theorem

Theorem 3.1.7. *Assume that, for some $\bar{t} \in [T_0, T_1]$,*

$$Span\left\{B_i(\bar{t})u \ : \ u \in \mathbb{R}^m, \ i \in \mathbb{N}\right\} = \mathbb{R}^n.$$

Then the linear control system $x' = A(t)x + B(t)u$ is controllable (on $[T_0, T_1]$).

Remark 3.1.8. *Our first remark is that the Cayley-Hamilton theorem can no longer be used: there are control systems $x' = A(t)x + B(t)u$ and $\bar{t} \in [T_0, T_1]$ such that*

$$Span\left\{B_i(\bar{t})u \ : \ u \in \mathbb{R}, \ i \in \mathbb{N}\right\} \neq Span\left\{B_i(\bar{t})u \ : \ u \in \mathbb{R}, \ i \in \{0, ..., n-1\}\right\}. \tag{3.1.11}$$

For example, let us take $T_0 = 0$, $T_1 = 1$, $n = m = 1$,$A(t) = 0$, $B(t) = t$. Then $B_0(t) = t$, $B_1(t) = 1$, $B_i(t) = 0$, $\forall i \in \mathbb{N} \setminus \{0, 1\}$. Therefore, if $\bar{t} = 0$, the left hand side of (3.1.11) *is $\mathbb{R}$ and the right hand side of* (3.1.11) *is $\{0\}$.*

Remark 3.1.9. *Our second remark is that the sufficient condition for controllability given in Theorem 3.1.7 is not a necessary condition (unless $n = 1$, or A and B are assumed to be analytic)*

3.1.2 Controllability of nonlinear systems in finite dimension

We consider the control system

$$\frac{dx}{dt} = f(x, u), \tag{3.1.12}$$

where $x \in \mathbb{R}^n$ is the state and $u \in \mathbb{R}^m$ is the control, with $(x, u) \in \Omega$ where Ω is a nonempty open subset of $\mathbb{R}^n \times \mathbb{R}^m$. Unless otherwise specified, we assume that $f \in \mathcal{C}^1(\Omega; \mathbb{R}^n)$. Let us give the definition of an equilibrium of the control system (3.1.12).

Definition 3.1.5. *An equilibrium of the control system* (3.1.12) *is a couple $(x_e, u_e) \in \Omega$ such that*

$$f(x_e, u_e) = 0.$$

Let us now give the definition we use for small-time local control- lability (we should in fact say small-time local controllability with controls close to u_e).

Definition 3.1.6. *Let $(x_e, u_e) \in \Omega$ be an equilibrium of the control system* (3.1.12)*. The control system* (3.1.12) *is small-time locally controllable at the equilibrium (x_e, u_e) if, for every real number $\varepsilon > 0$, there exists a real number $\eta > 0$ such that, for every*

$$x^0 \in B_\eta(x_e) := \{x \in \mathbb{R}^n \ : \ |x - x_e| < \eta\}$$

and for every $x^1 \in B_\eta(x_e)$, there exists a measurable function

$$u : [0, \varepsilon] \to \mathbb{R}^m$$

such that

$$|u(t) - u_e| \leq \varepsilon, \ \ \forall t \in [0, \varepsilon],$$
$$(x' = f(x, u(t)), \ x(0) = x^0) \Rightarrow (x(\varepsilon) = x^1).$$

One does not know any checkable necessary and sufficient condition for small-time local controllability for general control systems, even for analytic control systems. However, one knows powerful necessary conditions and powerful sufficient conditions. In this section we recall some of these conditions.

Now we prove that if a linearized control system at an equilibrium (resp. along a trajectory) is controllable, then the nonlinear control system is locally controllable at this equilibrium (resp. along this trajectory). We present an application of this useful result.

Local controllability along a trajectory and at an equilibrium point

We start with the definition of a trajectory.

Definition 3.1.7. *A trajectory of the control system* (3.1.12) *is a function* $(\overline{x},\overline{u}) : [T_0,T_1] \to \Omega$ *such that*

$$\overline{x} \in \mathcal{C}^0([T_0,T_1];\mathbb{R}^n), \quad \overline{u} \in L^\infty((T_0,T_1);\mathbb{R}^m),$$

exists a compact set $K \subset \Omega$ *such that* $(\overline{x}(t),\overline{u}(t)) \in K$, *for almost every* $t \in (T_0,T_1)$,

$$\overline{x}(t_2) = \overline{x}(t_1) + \int_{t_1}^{t_2} f(\overline{x}(t),\overline{u}(t))dt, \;\; \forall (t_1,t_2) \in [T_0,T_1]. \tag{3.1.13}$$

Let us make some comments on this definition:

- The definition 3.1.7 implies that

$$\frac{d\overline{x}}{dt} = f(\overline{x},\overline{u}) \;\text{ in } D'(T_0,T_1)^n. \tag{3.1.14}$$

  In (3.1.14), $D'(T_0,T_1)$ denotes the set of distributions on (T_0,T_1). Hence (3.1.14) just means that, for every $\varphi : (T_0,T_1) \to \mathbb{R}^n$ of class $\mathcal{C}^\infty$ and with compact support,

$$\int_{T_0}^{T_1} \left(\overline{x}(t)\cdot\varphi'(t) + f(\overline{x}(t),\overline{u}(t))\cdot\varphi(t)\right)dt = 0.$$

  For simplicity, when no confusion is possible, we write simply " $\dot{\overline{x}} = f(\overline{x},\overline{u})$ " instead of (3.1.14).

- If $\overline{u}$ is also continuous on $[T_0,T_1]$, then the condition that "exists a compact set $K \subset \Omega$ such that $(\overline{x}(t),\overline{u}(t)) \in K$, for almost every $t \in (T_0,T_1)$", is equivalent to

$$(\overline{x}(t),\overline{u}(t)) \in \Omega, \;\; \forall t \in [T_0,T_1].$$

  Moreover, in this case, $\overline{x}$ is of class $\mathcal{C}^1$ and (3.1.7) is equivalent to

$$\frac{d\overline{x}}{dt}(t) = f(\overline{x}(t),\overline{u}(t)), \;\; \forall t \in [T_0,T_1].$$

Let us now define the notion of "local controllability along a trajectory".

Definition 3.1.8. *Let* $(\overline{x}, \overline{u}) : [T_0, T_1] \to \Omega$ *be a trajectory of the control system* (3.1.12). *The control system* (3.1.12) *is locally controllable along the trajectory* $(\overline{x}, \overline{u})$ *if, for every* $\varepsilon > 0$, *there exists* $\eta > 0$ *such that, for every* $(a, b) \in \mathbb{R}^n \times \mathbb{R}^n$ *with* $|a - \overline{x}(T_0)| < \eta$ *and* $|b - \overline{x}(T_1)| < \eta$, *there exists a trajectory* $(x, u) : [T_0, T_1] \to \Omega$ *such that*

$$x(T_0) = a, \quad x(T_1) = b,$$
$$|u(t) - \overline{u}(t)| \leq \varepsilon, \quad \forall t \in [T_0, T_1].$$

Let us also introduce the definition of the linearized control system along a trajectory.

Definition 3.1.9. *The linearized control system along the trajectory* $(\overline{x}, \overline{u}) : [T_0, T_1] \to \Omega$ *is the linear time-varying control system*

$$\frac{dx}{dt} = \frac{\partial f}{\partial x}(\overline{x}(t), \overline{u}(t))x + \frac{\partial f}{\partial u}(\overline{x}(t), \overline{u}(t))u, \quad t \in [T_0, T_1],$$

where, at time $t \in [T_0, T_1]$, *the state is* $x(t) \in \mathbb{R}^n$ *and the control is* $u(t) \in \mathbb{R}^m$.

With these definitions, one can state the following classical and useful theorem.

Theorem 3.1.10. *Let* $(\overline{x}, \overline{u}) : [T_0, T_1] \to \Omega$ *be a trajectory of the control system* (3.1.12). *Let us assume that the linearized control system along the trajectory* $(\overline{x}, \overline{u})$ *is controllable (see Definition 3.1.2). Then the nonlinear control system* (3.1.12) *is locally controllable along the trajectory* $(\overline{x}, \overline{u})$.

Now let us consider $(x_e, u_e) \in \Omega$ an equilibrium point of the control system (3.1.12), and the constant trajectory $(\overline{x}, \overline{u})$ such that

$$(\overline{x}(t), \overline{u}(t)) = (x_e, u_e), \quad t \in [T_0, T_1].$$

Definition 3.1.10. *Let* (x_e, u_e) *be an equilibrium of the control system* (3.1.12). *The linearized control system at* (x_e, u_e) *of the control system* (3.1.12) *is the linear control system*

$$\frac{dx}{dt} = \frac{\partial f}{\partial x}(x_e, u_e)x + \frac{\partial f}{\partial u}(x_e, u_e)u, \quad t \in [T_0, T_1],$$

where, at time t, *the state is* $x(t) \in \mathbb{R}^n$ *and the control is* $u(t) \in \mathbb{R}^m$.

Then, from Theorem 3.1.10, we get the following classical theorem, which is a special case of a theorem due to Lawrence Markus [66].

Theorem 3.1.11. *Let* (x_e, u_e) *be an equilibrium point of the control system* (3.1.12). *Let us assume that the linearized control system of the control system* (3.1.12) *at* (x_e, u_e) *is controllable. Then the nonlinear control system* (3.1.12) *is small-time locally controllable at* (x_e, u_e).

Remark 3.1.12. *Theorem 3.1.11, which is easy and natural, is very useful. Let us recall that one can easily check whether the controllability of the linearized control system of the control system* (3.1.12) *at* (x_e, u_e) *holds by using the Kalman rank condition (Theorem 3.1.6).*

3.1.3 Return method

The return method has been introduced in [21] for a stabilization problem. It has been used for the first time in [22, 23] . In order to explain this method, let us first consider the problem of local controllability of a control system in finite dimension. Thus we consider the control system (3.1.12) and we assume that f is of class $\mathcal{C}^\infty$ and satisfies

$$f(0,0)=0.$$

The return method consists of reducing the local controllability of a nonlinear control system to the existence of (suitable) periodic trajectories and to the controllability of linear systems. The idea is the following one: Assume that, for every positive real number T and every positive real number ε, there exists a measurable bounded function $\overline{u}:[0,T]\to\mathbb{R}^m$ with $\|\overline{u}\|_{L^\infty(0,T)}\leq\varepsilon$ such that, if we denote by $\overline{x}$ the (maximal) solution of the Cauchy problem

$$\frac{d\overline{x}}{dt}=f(\overline{x},\overline{u}(t)),\ \ \overline{x}(0)=0,$$

then

$$\overline{x}(T)=0, \tag{3.1.15}$$

and the linearized control system around $(\overline{x},\overline{u})$ is controllable on $[0,T]$. Then, from Theorem 3.1.10, one gets the existence of $\eta>0$ such that, for every $x^0\in\mathbb{R}^n$ and for every $x^1\in\mathbb{R}^n$ such that

$$|x^0|<\eta,\ \ |x^1|<\eta,$$

there exists $u\in L^\infty((0,T);\mathbb{R}^m)$ such that

$$|u(t)-\overline{u}(t)|\leq\varepsilon,\ \ t\in[0,T],$$

and such that, if $x:[0,T]\to\mathbb{R}^n$ is the solution of the Cauchy problem

$$\frac{dx}{dt}(t)=f(x(t),u(t)),\ \ x(0)=x^0,$$

then

$$x(T)=x^1,$$

see Figure 3.1. Since $T>0$ and $\varepsilon>0$ are arbitrary, one gets that (3.1.12) is small-time locally controllable at the equilibrium $(0,0)\in\mathbb{R}^n\times\mathbb{R}^m$. (For the definition of small-time local controllability, see Definition 3.1.6).

Note that if one takes $\overline{u}=0$, then the above method just gives the well-known fact that, if the time-invariant linear control system

$$\frac{dy}{dt}=\frac{\partial f}{\partial x}(0,0)y+\frac{\partial f}{\partial u}(0,0)v,$$

is controllable, then the nonlinear control system (3.1.12) is small-time locally controllable at equilibrium point $(0,0)$ (Theorem 3.1.11. However, it may happen that the linearized control system around $(0,0)$, $(\overline{u}=0)$, is not controllable on $[0,T]$, but is controllable for other choices of $\overline{u}\neq 0)$.

h

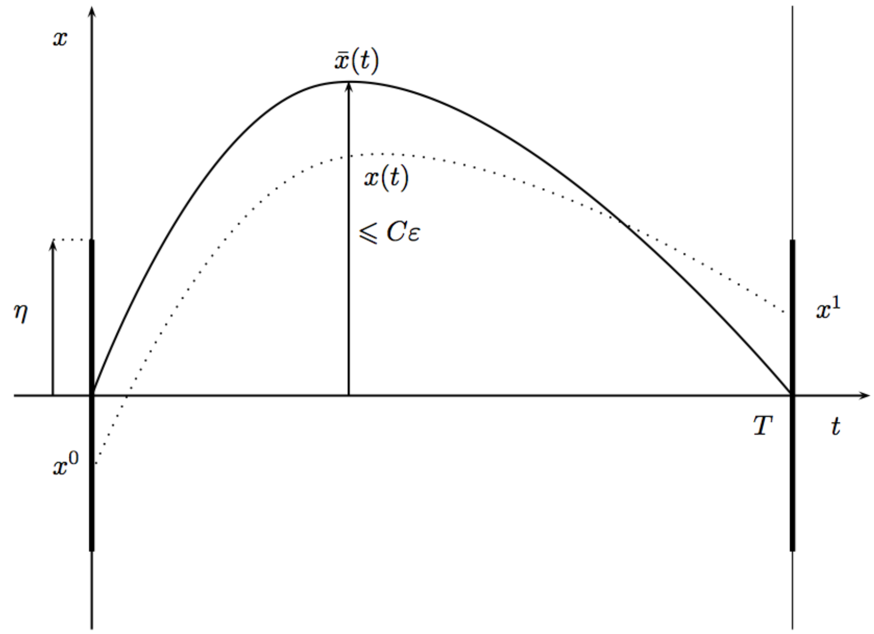


Figure 3.1: A sketch of the return method

3.2 Introduction

The control of boats or submarines has attracted the attention of the mathematical community from a long time (see e.g. [7, 9, 15, 34, 35, 60, 61, 62, 72].) In most of the papers devoted to that issue, the fluid is assumed to be inviscid, incompressible and irrotational, and the rigid body (the vehicle) is supposed to have an elliptic shape. On the other hand, to simplify the model, the control is often assumed to appear in a linear way in a finite-dimensional system describing the dynamics of the rigid body, the so-called *Kirchhoff laws*.

A large vessel (e.g. a cargo ship) presents often one tunnel thruster built into the bow to make docking easier. Some accurate model of a boat *without rudder* controlled by two propellers, the one displayed in a transversal bowthruster at the bow of the ship, the other one placed at the stern of the boat, was derived and investigated in [45]. A local controllability result was derived with two controls even if the fluid which could display some vorticity (i.e., the fluid was not assumed to be potential).

The aim of this paper is to provide some accurate model of an underwater vehicle immersed in an infinite volume of ideal fluid, without rudder, and actuated by a few number of propellers located into some tunnels inside the rigid body, and to give a rigorous analysis of the control properties of such a system. We aim to control both the position, the attitude, and the (linear and angular) velocities of the vehicle by taking as control input the flow of the fluid through a part of the boundary of the rigid body. The inviscid incompressible fluid is assumed here to be irrotational, hence potential, for the sake of simplicity. The case of a fluid with vorticity will be considered elsewhere.

Our fluid-structure interaction problem can be described as follow. The underwater vehicle, represented by a rigid body occupying a connected compact set $S(t) \subset \mathbb{R}^3$, is surrounded by an homogeneous incompressible perfect fluid filling the open set $\Omega(t) := \mathbb{R}^3 \setminus S(t)$ (as for e.g. a submarine immersed in an ocean). We assume that $\Omega(t)$ is C^∞ smooth and connected. Let $S = S(0)$ and

$$\Omega = \Omega(0) = \mathbb{R}^3 \setminus S(0)$$

denote the initial configuration ($t = 0$). Then, the dynamics of the fluid-structure system is governed by the following system of PDE's

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p = 0, \qquad t \in (0,T),\ x \in \Omega(t) \tag{3.2.1}$$

$$\text{div } u = 0, \qquad t \in (0,T),\ x \in \Omega(t) \tag{3.2.2}$$

$$u \cdot \hat{n} = (h' + \omega \times (x - h)) \cdot \hat{n} + w(t,x), \qquad t \in (0,T),\ x \in \partial\Omega(t) \tag{3.2.3}$$

$$\lim_{|x| \to +\infty} u(t,x) = 0, \tag{3.2.4}$$

$$mh'' = \int\limits_{\partial\Omega(t)} p\hat{n}\, d\sigma, \qquad t \in (0,T) \tag{3.2.5}$$

$$\frac{d}{dt}(QJ_0Q^*\omega) = \int\limits_{\partial\Omega(t)} (x - h) \times p\hat{n}\, d\sigma, \qquad t \in (0,T) \tag{3.2.6}$$

$$Q' = S(\omega)Q, \qquad t \in (0,T) \tag{3.2.7}$$

$$u(0,x) = u_0(x), \qquad x \in \Omega, \tag{3.2.8}$$

$$(h(0), Q(0), h'(0), \omega(0)) = (h_0, Q_0, h_1, \omega_0) \in \mathbb{R}^3 \times \mathrm{SO}(3) \times \mathbb{R}^6. \tag{3.2.9}$$

In the above equations, u (resp. p) is the velocity field (resp. the pressure) of the fluid, h denotes the position of the center of mass of the solid, ω denotes the angular velocity and Q the 3 dimensional rotation matrix giving the orientation of the solid. The positive constant m and the matrix J_0, which denote respectively the mass and the inertia matrix of the rigid body, are defined as

$$m = \int\limits_S \rho(x)dx, \quad J_0 = \int\limits_S \rho(x)(|x|^2 Id - xx^*)dx,$$

where $\rho(\cdot)$ represents the density of the rigid body. $\hat{n}$ is the outward unit vector to $\partial\Omega(t)$, $x \times y$ is the cross product between the vectors x and y, and $S(y)$ is the skew-adjoint matrix such that $S(y)x = y \times x$, i.e.

$$S(y) = \begin{pmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{pmatrix}.$$

$\dot{f}$ stands for the derivative of f respect to t, A^* means the transpose of the matrix A, and Id denotes the matrix identity. Finally, the term $w(t,x)$, which stands for the flow through the boundary of the rigid body, is taken as control input. Its support will be strictly included in $\partial\Omega(t)$, and actually only a finite dimensional control input will be considered here (see below (3.2.16) for the precise form of the control term $w(t,x)$).

When no control is applied (i.e. $w(t,x)=0$), then the existence and uniqueness of strong solutions to (3.2.1)-(3.2.9) was obtained first in [73] for a ball in dimension 2, and next in [74] for a rigid body S of arbitrary form. The result in [73] was extended to any dimension in [80] (in that paper, the issue of the persistence of regularity is also studied). We also refer to [46] for the situation when $\Omega(t)=\Omega_0\setminus\overline{S(t)}$, with Ω_0 a bounded open set in $\mathbb{R}^3$, and for the issue of the analyticity in time. The detection of the rigid body $S(t)$ from partial measurements of the fluid velocity has been tackled in [19] when $\Omega(t)=\Omega_0\setminus\overline{S(t)}$ ($\Omega_0\subset\mathbb{R}^2$ still being a bounded cavity) and in [20] when $\Omega(t)=\mathbb{R}^2\setminus\overline{S(t)}$.

Here, we are interested in the control properties of (3.2.1)-(3.2.9). The controllability of Euler equations has been established in 2D (resp. in 3D) in [23] (resp. in [42]). Note, however, that there is no hope here to control both the fluid and the rigid body motion. Indeed, $\Omega(t)$ is an exterior domain, and the vorticity is transported by the flow with a finite speed propagation, so that it is not affected (at any given time) far from the boat. Therefore, we will deal with the control of the motion of the rigid body only. As the state of the rigid body is described by a vector in $\mathbb{R}^{12}$, it is natural to consider a finite-dimensional control input.

Note also that since the fluid is flowing through a part of the boundary of the rigid body, one more boundary condition is needed to ensure the uniqueness of the solution of (3.2.1)-(3.2.9) (see [49], [51]). In dimension three, one can impose the value of the vorticity $\zeta(t,x):=\text{curl }v(t,x)$ on the inflow section; that is, we may set

$$\zeta(t,x)=\zeta_0(t,x)\ \text{ for } w(t,x)<0 \tag{3.2.10}$$

where $\zeta_0(t,x)$ is a given function.

In order to write the equation of the fluid in a *fixed frame*, we perform a change of coordinates. We set

$$\begin{aligned}
x &= Q(t)y+h(t), && (3.2.11)\\
v(t,y) &= Q^*(t)u(t,Q(t)y+h(t)), && (3.2.12)\\
q(t,y) &= p(t,Q(t)y+h(t)), && (3.2.13)\\
l(t) &= Q^*(t)h'(t), && (3.2.14)\\
r(t) &= Q^*(t)\omega(t). && (3.2.15)
\end{aligned}$$

Then x (resp. y) represents the vector of coordinates of a point in a fixed frame (respectively in a frame linked to the rigid body). Note that, at any given time t, y ranges over the fixed domain Ω when x ranges over $\Omega(t)$. Finally, we assume that the control takes the form

$$w(t,x)=w(t,Q(t)y+h(t))=\sum_{j=1}^{m}w_j(t)\chi_j(y), \tag{3.2.16}$$

where $n\in\mathbb{N}^*$ stands for the number of independent inputs, and $w_j(t)\in\mathbb{R}$ is the control input associated with the function $\chi_j \in C^\infty(\partial\Omega)$. To ensure the conservation of the mass of the fluid, we impose the relation

$$\int_{\partial\Omega} \chi_j(y) d\sigma = 0 \quad \text{for } 1 \le j \le m. \tag{3.2.17}$$

Then the functions (v, q, l, r) satisfy the following system

$$\frac{\partial v}{\partial t} + ((v - l - r \times y) \cdot \nabla)v + r \times v + \nabla q = 0, \qquad t \in (0,T),\ y \in \Omega, \tag{3.2.18}$$

$$\operatorname{div} v = 0, \qquad t \in (0,T),\ y \in \Omega, \tag{3.2.19}$$

$$v \cdot \hat{n} = (l + r \times y) \cdot \hat{n} + \sum_{1 \le j \le m} w_j(t)\chi_j(y), \qquad t \in (0,T),\ y \in \partial\Omega, \tag{3.2.20}$$

$$\lim_{|y| \to +\infty} v(t, y) = 0, \tag{3.2.21}$$

$$m\dot{l} = \int_{\partial\Omega} q\hat{n}\, d\sigma - mr \times l, \qquad t \in (0,T), \tag{3.2.22}$$

$$J_0\dot{r} = \int_{\partial\Omega} q(y \times \hat{n})\, d\sigma - r \times J_0 r, \qquad t \in (0,T), \tag{3.2.23}$$

$$(l(0), r(0)) = (h_1, r_0),\ v(0, y) = u_0(y). \tag{3.2.24}$$

The work is organized as follows. In Section 3.3, we simplify system (3.2.1)-(3.2.9) by assuming that the fluid is potential. We obtain a finite dimensional system (namely (3.3.64)) similar to Kirchhoff laws, in which the control input w appears through both linear terms (with time derivative) and bilinear terms. The investigation of the control properties of (3.3.64) is performed in Section 3.4. After noticing that the controllability of the linearized system at the origin requires six control inputs, we apply the *return method* due to Jean-Michel Coron to take advantage of the nonlinear terms in (3.3.64). (We refer the reader to [24] for an exposition of that method for finite-dimensional systems and for PDEs.) We consider the linearization along a certain closed-loop trajectory and obtain a local controllability result (Theorem 3.4.9) assuming that two rank conditions are fulfilled, by using a variant of Silverman-Meadows test for the controllability of a time-varying linear system. Some examples using symmetry properties of the rigid body are given in Section 3.5.

3.3 Equations of the motion in the potential case

In this section we derive the equations describing the motion of the rigid body subject to flow boundary control when the fluid is potential.

3.3.1 Null vorticity

Let us denote by

$$\zeta(t,y) = \operatorname{rot} v(t,y) := (\nabla \times v)(t,y)$$

the *vorticity* of the fluid. Here, we assume that

$$\zeta_0 = \operatorname{rot} v_0 = 0 \quad \text{in } \Omega \tag{3.3.1}$$

and that ζ is null at the inflow part of $\partial\Omega$; that is,

$$\zeta(t,y) = 0 \quad \text{if} \;\; w_j(t)\chi_j(y) < 0, \quad \text{for some } j \in [1,n]. \tag{3.3.2}$$

Proposition 3.3.1. *Under the assumptions (3.3.1) and (3.3.2), one has*

$$\zeta = \operatorname{rot} v \equiv 0 \quad \textit{in } (0,T) \times \Omega, \tag{3.3.3}$$

Proof. Let us introduce $\tilde{v} := v - l - r \times y$. Then it follows from (3.2.19) that

$$\operatorname{div}(\tilde{v}) = 0, \tag{3.3.4}$$

and

$$\operatorname{rot}(\tilde{v}) = \omega - 2r. \tag{3.3.5}$$

Applying the operator rot in (3.2.18) results in

$$\frac{\partial \zeta}{\partial t} + \operatorname{rot}((v\cdot\nabla)v) - \operatorname{rot}((l + r \times y)\cdot\nabla v) = 0 \tag{3.3.6}$$

We note that the following identities hold:

$$\operatorname{rot}((v\cdot\nabla)v) = (v\cdot\nabla)\operatorname{rot}(v) - (\operatorname{rot}(v)\cdot\nabla)v + \operatorname{div}(v)\operatorname{rot}(v) \tag{3.3.7}$$

and

$$\operatorname{rot}(r \times v) = \operatorname{div}(v)r - (r\cdot\nabla)v. \tag{3.3.8}$$

Using (3.3.4)-(3.3.8), we obtain

$$\frac{\partial \zeta}{\partial t} + (\tilde{v}\cdot\nabla)\zeta - (\zeta\cdot\nabla)\tilde{v} = 0. \tag{3.3.9}$$

Let $\varphi = \varphi(t,s,y)$ denote the flow associated with $\tilde{v}$, i.e.

$$\frac{\partial \varphi}{\partial t} = \tilde{v}(t,\varphi), \;\text{with}\; \varphi|_{t=s} = y. \tag{3.3.10}$$

We denote by $G(t,s,y) = \frac{\partial\varphi}{\partial y}(t,s,y)$ the Jacobi matrix of φ. Differentiating in (3.3.10) with respect to y_j $(j = 1,2,3)$, we see that $G(t,s,y)$ satisfies the following equation:

$$\frac{\partial G}{\partial t} = \frac{\partial \tilde{v}}{\partial y}(t,\varphi(t,s,y))\cdot G(t,s,y), \;\text{where}\;\; G(s,s,y) = Id \text{ (identity matrix)}. \tag{3.3.11}$$

We infer from (3.3.4) and (3.3.11) that

$$\det G(t,s,y) = 1. \tag{3.3.12}$$

Following Yudovich [49], we introduce the time $t^*(t,y)$ at which the fluid element first appears in the domain, and set $y^*(t,y) = \varphi(t^*(t,y),t,y)$. Then either $t^* = 0$, or $t^* > 0$ and $y^* \in \partial\Omega$ with $w_j(t^*)\chi_j(y^*) < 0$ for some $j \in [1,n]$. Set $f(s,t,y) = G^{-1}(s,t,y)\zeta(s,\varphi(s,t,y))$. From (3.3.9)-(3.3.12), we obtain that

$$\frac{\partial f}{\partial s}(s,t,y) = 0. \tag{3.3.13}$$

Finally, integrating with respect to s in (3.3.13) yields

$$\zeta(t,y) = G^{-1}(t^*,t,y)\zeta(t^*,y^*), \tag{3.3.14}$$

which, combined to (3.3.1) and (3.3.2), gives (3.3.3). The proof of Proposition 3.3.1 is complete. □

3.3.2 Decomposition of the fluid velocity

It follows from (3.2.19), (3.2.21) and (3.3.3) that the flow is potential; that is,

$$v = \nabla\Phi, \tag{3.3.15}$$

where $\Phi = \Phi(t,y)$ solves

$$\Delta\Phi = 0, \quad \text{in } (0,T)\times\Omega, \tag{3.3.16}$$

$$\frac{\partial\Phi}{\partial n} = (l + r\times y)\cdot\hat{n} + \sum_{1\le j\le n} w_j(t)\chi_j(y) \quad \text{on } (0,T)\times\Omega, \tag{3.3.17}$$

$$\lim_{|y|\to+\infty} \nabla\Phi(t,y) = 0, \quad \text{on } (0,T). \tag{3.3.18}$$

Actually, Φ may be decomposed as

$$\Phi(t,y) = \sum_{1\le i\le 3}\left\{l_i\phi_i + r_i\varphi_i\right\} + \sum_{1\le j\le n} w_j\psi_j \tag{3.3.19}$$

where, for $i = 1,2,3$ and $j = 1,...,n$,

$$\Delta\phi_i = \Delta\varphi_i = \Delta\psi_j = 0 \text{ in } \Omega, \tag{3.3.20}$$

$$\frac{\partial\phi_i}{\partial\hat{n}} = \hat{n}_i, \quad \frac{\partial\varphi_i}{\partial\hat{n}} = (y\times\hat{n})_i, \quad \frac{\partial\psi_j}{\partial\hat{n}} = \chi_j \text{ on } \partial\Omega, \tag{3.3.21}$$

$$\lim_{|y|\to+\infty}\nabla\phi_i(y) = 0, \quad \lim_{|y|\to+\infty}\nabla\varphi_i(y) = 0, \quad \lim_{|y|\to+\infty}\nabla\psi_j(y) = 0, \tag{3.3.22}$$

As the domain S occupied by the rigid body and the functions χ_j, $1\le j\le n$, supporting the control are assumed to be smooth, we infer that the functions $\nabla\phi_i$ ($i = 1,2,3$), the functions $\nabla\varphi_i$ ($i = 1,2,3$) and the functions $\nabla\psi_j$ ($1\le j\le n$) belong to $H^\infty(\Omega)$.

3.3.3 Equations for the linear and angular velocities

Let us introduce the matrices $M, J, D \in \mathbb{R}^{3\times3}$, $C^M, C^J \in \mathbb{R}^{3\times n}$, $L_p^M, L_p^J, R_p^M, R_p^J \in \mathbb{R}^{3\times3}$, and the matrices $W_p^M, W_p^J \in \mathbb{R}^{3\times n}$ for $p \in [1,..,n]$, defined by

$$M_{i,j} = \int_{\partial\Omega} \hat{n}_i \phi_j = \int_{\partial\Omega} \frac{\partial \phi_i}{\partial \hat{n}} \phi_j = \int_{\Omega} \nabla\phi_i \cdot \nabla\phi_j, \tag{3.3.23}$$

$$J_{i,j} = \int_{\partial\Omega} (y \times \hat{n})_i \varphi_j = \int_{\partial\Omega} \frac{\partial \varphi_i}{\partial \hat{n}} \varphi_j = \int_{\Omega} \nabla\varphi_i \cdot \nabla\varphi_j, \tag{3.3.24}$$

$$D_{i,j} = \int_{\Omega} \nabla\phi_i \cdot \nabla\varphi_j = \int_{\partial\Omega} \hat{n}_i \varphi_j = \int_{\partial\Omega} \phi_i (y \times \hat{n})_j, \tag{3.3.25}$$

$$(C^M)_{i,j} = \int_{\Omega} \nabla\phi_i \cdot \nabla\psi_j = \int_{\partial\Omega} \hat{n}_i \psi_j = \int_{\partial\Omega} \phi_i \chi_j, \tag{3.3.26}$$

$$(C^J)_{i,j} = \int_{\Omega} \nabla\varphi_i \cdot \nabla\psi_j = \int_{\partial\Omega} (y \times \hat{n})_i \psi_j = \int_{\partial\Omega} \varphi_i \chi_j, \tag{3.3.27}$$

$$(L_p^M)_{i,j} = \int_{\partial\Omega} (\nabla\phi_j)_i \chi_p, \qquad (L_p^J)_{i,j} = \int_{\partial\Omega} (y \times \nabla\phi_j)_i \chi_p, \tag{3.3.28}$$

$$(R_p^M)_{i,j} = \int_{\partial\Omega} (\nabla\varphi_j)_i \chi_p, \qquad (R_p^J)_{i,j} = \int_{\partial\Omega} (y \times \nabla\varphi_j)_i \chi_p, \tag{3.3.29}$$

$$(W_p^M)_{i,j} = \int_{\partial\Omega} (\nabla\psi_j)_i \chi_p, \qquad (W_p^J)_{i,j} = \int_{\partial\Omega} (y \times \nabla\psi_j)_i \chi_p. \tag{3.3.30}$$

Note that $M^* = M$ and $J^* = J$.

Let us now reformulate the equations for the motion of the rigid body. We define the matrix $\mathcal{J} \in \mathbb{R}^{6\times6}$ by

$$\mathcal{J} = \begin{pmatrix} mI_d & 0 \\ 0 & J_0 \end{pmatrix} + \begin{pmatrix} M & D \\ D^* & J \end{pmatrix}. \tag{3.3.31}$$

It is easy to see that $\mathcal{J}$ is a (symmetric) positive definite matrix. We associate to the (linear and angular) velocity $(l, r) \in \mathbb{R}^3 \times \mathbb{R}^3$ of the rigid body a momentum-like quantity, the so-called *impulse* $(P, \Pi) \in \mathbb{R}^3 \times \mathbb{R}^3$, defined by

$$\mathcal{J} \begin{pmatrix} l \\ r \end{pmatrix} = \begin{pmatrix} P \\ \Pi \end{pmatrix}. \tag{3.3.32}$$

We are now in a position to give the equations governing the dynamics of the impulse.

Proposition 3.3.2. *The dynamics of the system are governed by the following Kirchhoff equations*

$$\begin{aligned}\frac{dP}{dt}+C^M\dot{w} &= (P+C^Mw)\times r-\sum_{1\le p\le n} w_p\left\{L_p^M l+R_p^M r+W_p^M w\right\},\\ \frac{d\Pi}{dt}+C^J\dot{w} &= (\Pi+C^Jw)\times r+P\times l-\sum_{1\le p\le n} w_p\left\{L_p^J l+R_p^J r+W_p^J w\right\},\end{aligned}\tag{3.3.33}$$

where $w(t):=(w_1(t),...,w_n(t))\in\mathbb{R}^n$ *denotes the control input.*

Proof. We first express the pressure q in terms of l,r,v and their derivatives. Using (3.3.3), we easily obtain

$$v\cdot\nabla v=\nabla\frac{|v|^2}{2}\quad\text{and}\quad (r\times y)\cdot\nabla v-r\times v=\nabla((r\times y)\cdot v)\tag{3.3.34}$$

Thus (3.2.18) gives

$$\begin{aligned}-\nabla q &= \frac{\partial v}{\partial t}+\nabla\left(\frac{|v|^2}{2}-l\cdot v-(r\times y)\cdot v\right)\\ &= \nabla\left(\sum_{1\le i\le 3}\left\{l_i'\phi_i+r_i'\varphi_i\right\}+\sum_{1\le j\le n}w_j'\psi_j+\frac{|v|^2}{2}-l\cdot v-(r\times y)\cdot v\right)\end{aligned}$$

hence we can take

$$q=-\left\{\sum_{1\le i\le 3}\left\{l_i'\phi_i+r_i'\varphi_i\right\}+\sum_{1\le j\le n}w_j'\psi_j+\frac{|v|^2}{2}-(l+(r\times y))\cdot v\right\}\tag{3.3.35}$$

Replacing q by its value in (3.2.22) yields

$$m\dot{l}=-mr\times l-\left\{\sum_{1\le i\le 3}\left\{l_i'\int_{\partial\Omega}\phi_i\hat{n}+r_i'\int_{\partial\Omega}\varphi_i\hat{n}\right\}+\sum_{1\le j\le n}w_j'\int_{\partial\Omega}\psi_j\hat{n}+\int_{\partial\Omega}\left(\frac{|v|^2}{2}-(l+(r\times y))\cdot v\right)\hat{n}\right\}.\tag{3.3.36}$$

Using (3.3.34), (3.2.19-3.2.20), we obtain

$$\begin{aligned}\int_{\partial\Omega}\frac{|v|^2}{2}\nu &= \int_{\Omega}\nabla\frac{|v|^2}{2}\\ &= \int_{\Omega}v\cdot\nabla v\\ &= -\int_{\Omega}(\operatorname{div}v)v+\int_{\partial\Omega}(v\cdot\hat{n})v\\ &= \int_{\partial\Omega}((l+r\times y)\cdot\hat{n})\,v+\sum_{1\le j\le n}w_j(t)\int_{\partial\Omega}\chi_j(y)v.\end{aligned}\tag{3.3.37}$$

Using the following identity

$$a \times (v \times \hat{n}) = (a \cdot \hat{n})v - (a \cdot v)\hat{n}, \quad \forall v, a \in \mathbb{R}^3, \tag{3.3.38}$$

we obtain that

$$\int\limits_{\partial\Omega} ((l + r \times y) \cdot \hat{n})\, v - ((l + (r \times y)) \cdot v)\, \hat{n} = \int\limits_{\partial\Omega} (l + r \times y) \times (v \times \hat{n}). \tag{3.3.39}$$

Now we claim that

$$\int\limits_{\partial\Omega} \hat{n} \times \nabla f = 0, \quad \forall f \in H^2(\Omega). \tag{3.3.40}$$

To prove the claim, we introduce a smooth cutoff function ρ_a such that

$$\rho_a(y) = \begin{cases} 1 & \text{if } |y| < a, \\ 0 & \text{if } |y| > 2a. \end{cases}$$

Pick a radius $a > 0$ such that $S(0) \subset B(0, a)$, and set

$$\tilde{f}(y) = f(y)\rho_a(y). \tag{3.3.41}$$

Then

$$\nabla \tilde{f}(y) = \nabla f(y), \quad \forall y \in \partial\Omega,$$

and using the divergence Theorem, we obtain

$$\int\limits_{\partial\Omega} \hat{n} \times \nabla f = \int\limits_{\partial\Omega} \hat{n} \times \nabla \tilde{f} = \int\limits_{\Omega} \text{rot } (\nabla \tilde{f}) = 0.$$

Therefore, using (3.3.40) with $f = \Phi$ where $\nabla\Phi = v$, we obtain

$$\int\limits_{\partial\Omega} l \times (v \times \hat{n}) = 0. \tag{3.3.42}$$

Another application of (3.3.40) with $f = y_i\Phi$ yields

$$\int\limits_{\partial\Omega} y_i v \times \hat{n} = \int\limits_{\partial\Omega} \hat{n} \times \hat{e}_i \Phi. \tag{3.3.43}$$

Combined to (3.3.39), this gives

$$\int\limits_{\partial\Omega} (l + r \times y) \times (v \times \hat{n}) = r \times \int\limits_{\partial\Omega} \Phi\hat{n}$$

$$= r(t) \times \left(\sum_{i=1}^{3} \left\{ l_i(t) \int\limits_{\partial\Omega} \phi_i(y)\hat{n}(y) + r_i(t) \int\limits_{\partial\Omega} \varphi_i(y)\hat{n}(y) \right\} + \sum_{j=1}^{m} w_j(t) \int\limits_{\partial\Omega} \psi_j(y)\hat{n}(y) \right). \tag{3.3.44}$$

Replacing (3.3.37) and (3.3.44) in (3.3.36) results in

$$\begin{aligned} m\dot{l} &= -\left\{\sum_{i=1}^{3} \dot{l}_i \int\limits_{\partial\Omega} \phi_i \hat{n} + \dot{r}_i \int\limits_{\partial\Omega} \varphi_i \hat{n} + \sum_{j=1}^{n} \dot{w}_j \int\limits_{\partial\Omega} \psi_j \hat{n}\right\} \\ &\quad -\sum_{j=1}^{n} w_j \left\{\sum_{i=1}^{3} l_i \int\limits_{\partial\Omega} \chi_j \nabla\phi_i + r_i \int\limits_{\partial\Omega} \chi_j \nabla\varphi_i + \sum_{p=1}^{n} w_p \int\limits_{\partial\Omega} \chi_j \nabla\psi_p\right\} \\ &\quad -r \times \left\{\sum_{i=1}^{3} l_i \int\limits_{\partial\Omega} \phi_i \hat{n} + r_i \int\limits_{\partial\Omega} \varphi_i \hat{n} + \sum_{j=1}^{n} w_j \int\limits_{\partial\Omega} \psi_j \hat{n}\right\} \\ &\quad -mr \times l. \end{aligned} \tag{3.3.45}$$

Let us turn our attention to the dynamics of r. Substituting the expression of q given in (3.3.35) in (3.2.23) yields

$$\begin{aligned} J_0\dot{r} = -r \times J_0 r - \sum_{1\le i\le 3} \left\{ l_i' \int\limits_{\partial\Omega} \phi_i(y \times \hat{n}) + r_i' \int\limits_{\partial\Omega} \varphi_i(y \times \hat{n}) \right\} - \sum_{1\le j\le n} w_j' \int\limits_{\partial\Omega} \psi_j(y \times \hat{n}) \\ - \int\limits_{\partial\Omega} \left(\frac{|v|^2}{2} - (l + (r \times y)) \cdot v \right) (y \times \hat{n}) \end{aligned} \tag{3.3.46}$$

Using (3.3.34), (3.2.19) and (3.2.20), we obtain that

$$\begin{aligned} \int_{\partial\Omega} \left(\frac{|v|^2}{2}\right)(y \times \hat{n})_i &= \int\limits_{\Omega} \operatorname{div}\left(\left(\frac{|v|^2}{2}\right)(\hat{e}_i \times y)\right) \\ &= \int\limits_{\partial\Omega} (v \cdot \hat{n})(y \times v)_i - \int\limits_{\Omega} \operatorname{div}(v)\,(y \times v)_i - \int\limits_{\Omega} (v \times v)_i \\ &= \int\limits_{\partial\Omega} (v \cdot \hat{n})(y \times v)_i \\ &= \int\limits_{\partial\Omega} (l + r \times y) \cdot \hat{n}(y \times v)_i + \sum_{1\le j\le n} w_j(t) \int\limits_{\partial\Omega} \chi_j (y \times v)_i. \end{aligned} \tag{3.3.47}$$

Furthermore, using (3.3.38) we have

$$\begin{aligned}
&\int_{\partial\Omega} (l + r \times y) \cdot \hat{n}(y \times v)_i - (l + (r \times y)) \cdot v(y \times \hat{n})_i \\
&\quad = \int_{\partial\Omega} (l + r \times y) \cdot (\hat{n}(y \times v)_i - v(y \times \hat{n})_i) \\
&\quad = \int_{\partial\Omega} (l + r \times y) \cdot (\hat{n}((\hat{e}_i \times y) \cdot v) - v((\hat{e}_i \times y) \cdot \hat{n})) \\
&\quad = \int_{\partial\Omega} (l + r \times y) \cdot ((\hat{e}_i \times y) \times (\hat{n} \times v)) \,.
\end{aligned} \tag{3.3.48}$$

Combining the following identity

$$\sum_{j=1}^{3} (a \times \hat{e}_j) \times (\hat{e}_j \times b) = -(a \times b), \quad \forall a, b \in \mathbb{R}^3 \tag{3.3.49}$$

to (3.3.43), we obtain

$$\begin{aligned}
\int_{\partial\Omega} l \cdot ((\hat{e}_i \times y) \times (\hat{n} \times v)) &= \sum_{j=1}^{3} \int_{\partial\Omega} l \cdot ((\hat{e}_i \times \hat{e}_j) \times (\hat{n} \times y_j v)) \\
&= \sum_{j=1}^{3} l \cdot \left((\hat{e}_i \times \hat{e}_j) \times \Big(\int_{\partial\Omega} (\hat{n} \times y_j v) \Big) \right) \\
&= \sum_{j=1}^{3} l \cdot \left((\hat{e}_i \times \hat{e}_j) \times \Big(\int_{\partial\Omega} (\hat{e}_j \times \hat{n}) \Phi \Big) \right) \\
&= -l \cdot \int_{\partial\Omega} (\hat{e}_i \times \hat{n})\, \Phi = \int_{\partial\Omega} (l \times \hat{n})_i\, \Phi.
\end{aligned} \tag{3.3.50}$$

Let

$$I = \int_{\partial\Omega} (r \times y) \cdot ((\hat{e}_i \times y) \times (\hat{n} \times \nabla f)) \,.$$

$\tilde{f}$ denoting still the function defined in (3.3.41), we have that

$$\begin{aligned}
I &= \sum_{j=1}^{3} \int_{\partial\Omega} (r \times y) \cdot \Big((\hat{e}_i \times y) \times (\hat{e}_j \times \nabla \tilde{f}) \Big) n_j \\
&= \sum_{j=1}^{3} \left\{ \int_{\Omega} (r \times \hat{e}_j) \cdot \Big((\hat{e}_i \times y) \times (\hat{e}_j \times \nabla \tilde{f}) \Big) \right\} + \sum_{j=1}^{3} \left\{ \int_{\Omega} (r \times y) \cdot \Big((\hat{e}_i \times \hat{e}_j) \times (\hat{e}_j \times \nabla \tilde{f}) \Big) \right\} \\
&\qquad + \sum_{j=1}^{3} \left\{ \int_{\Omega} (r \times y) \cdot \Big((\hat{e}_i \times y) \times (\hat{e}_j \times \partial_j \nabla \tilde{f}) \Big) \right\}
\end{aligned}$$

Using again (3.3.49), we obtain

$$\begin{aligned}
I &= -\sum_{j=1}^{3}\left\{\int_{\Omega}(\hat{e}_i\times y)\cdot\Big((r\times\hat{e}_j)\times(\hat{e}_j\times\nabla\tilde{f})\Big)\right\}-\int_{\Omega}(r\times y)\cdot(\hat{e}_i\times\nabla\tilde{f}) \\
&\quad +\int_{\Omega}(r\times y)\cdot\Big((\hat{e}_i\times y)\times\mathrm{rot}(\nabla\tilde{f})\Big) \\
&= \int_{\Omega}(\hat{e}_i\times y)\cdot(r\times\nabla\tilde{f})-\int_{\Omega}(r\times y)\cdot(\hat{e}_i\times\nabla\tilde{f}) \\
&= -\int_{\Omega}r\cdot(\hat{e}_i\times y)\times\nabla\tilde{f}-\int_{\Omega}r\cdot y\times(\hat{e}_i\times\nabla\tilde{f}) \\
&= -\int_{\Omega}r\cdot\Big\{(\hat{e}_i\times y)\times\nabla\tilde{f}+y\times(\hat{e}_i\times\nabla\tilde{f})\Big\} \\
&= -\int_{\Omega}r\cdot\Big\{\hat{e}_i\times(y\times\nabla\tilde{f})\Big\}=\int_{\Omega}\Big(r\times(y\times\nabla\tilde{f})\Big)_i=\int_{\partial\Omega}(r\times(y\times\hat{n})f)_i\,.
\end{aligned}$$

Letting $f=\Phi$ in the above expression yields

$$\int_{\partial\Omega}(r\times y)\cdot((\hat{e}_i\times y)\times(\hat{n}\times v))=\int_{\partial\Omega}(r\times(y\times\hat{n})\Phi)_i\,. \tag{3.3.51}$$

Therefore, replacing (3.3.47,3.3.48,3.3.50) and (3.3.51) in (3.3.46) yields

$$\begin{aligned}
J_0\dot{r} &= \sum_{i=1}^{3}\Big(\dot{l}_i\int_{\partial\Omega}(\hat{n}\times y)\phi_i+\dot{r}_i\int_{\partial\Omega}(\hat{n}\times y)\varphi_i\Big)+\sum_{j=1}^{n}\dot{w}_j\int_{\partial\Omega}(\hat{n}\times y)\psi_j \\
&+ \sum_{j=1}^{n}w_j\left\{\sum_{i=1}^{3}\Big(l_i\int_{\partial\Omega}(\nabla\phi_i\times y)\chi_j+r_i\int_{\partial\Omega}(\nabla\varphi_i\times y)\chi_j\Big)+\sum_{p=1}^{n}w_p\int_{\partial\Omega}(\nabla\psi_p\times y)\chi_j\right\} \\
&- l\times\left\{\sum_{i=1}^{3}\Big(l_i\int_{\partial\Omega}\phi_i\hat{n}+r_i\int_{\partial\Omega}\varphi_i\hat{n}\Big)+\sum_{p=1}^{n}w_p\int_{\partial\Omega}\psi_p\hat{n}\right\} \\
&- r\times\left\{\sum_{i=1}^{3}\Big(l_i\int_{\partial\Omega}(y\times\hat{n})\phi_i+r_i\int_{\partial\Omega}(y\times\hat{n})\varphi_i\Big)+\sum_{p=1}^{n}w_p\int_{\partial\Omega}(y\times\hat{n})\psi_p\right\} \\
&- r\times J_0 r.
\end{aligned} \tag{3.3.52}$$

Combining (3.3.45) and (3.3.52) to the definitions of the matrices in (3.3.23)-(3.3.30), we obtain

$$
\begin{aligned}
m\dot{l} &= -M\dot{l} - D\dot{r} - C^M\dot{w} - r \times (Ml + Dr + C^M w) \\
&\quad -\sum_{p=1}^{n} w_p \left\{ L_p^M l + R_p^M r + W_p^M w \right\} - mr \times l, \qquad (3.3.53) \\
J_0\dot{r} &= -D^*\dot{l} - J^R\dot{r} - C^J\dot{w} - r \times (D^* l + J^R r + C^J w) \\
&\quad -\sum_{p=1}^{n} w_p \left\{ L_p^J l + R_p^J r + W_p^J w \right\} \\
&\quad -l \times (Ml + Dr + C^M w) - r \times J_0 r. \qquad (3.3.54)
\end{aligned}
$$

This completes the proof of Proposition 3.3.2. □

3.3.4 Equations for the position and attitude

Now, we look at the dynamics of the position and attitude. We shall use unit quaternions. (We refer the reader to the Section 3.6 for the notations and definitions used in what follows). From (3.2.7), we obtain

$$Q' = S(Qr)Q = QS(r) \qquad (3.3.55)$$

where $Q(0) = Id$.

Assuming that $Q(t)$ is associated with a unit quaternion $q(t)$, i.e. $Q(t) = R(q(t))$, then the dynamics of q is given by

$$\dot{q} = \frac{1}{2} q * r \qquad (3.3.56)$$

(see e.g. [88]). Expanding q as $q = q_0 + \vec{q} = q_0 + q_1 i + q_2 j + q_3 k$, this yields

$$\dot{q}_0 + \dot{\vec{q}} = \frac{1}{2}(-\vec{q} \cdot r + q_0 r + \vec{q} \times r) \qquad (3.3.57)$$

and

$$\begin{pmatrix} \dot{q}_0 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{pmatrix} = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} 0 \\ r_1 \\ r_2 \\ r_3 \end{pmatrix}. \qquad (3.3.58)$$

From (3.2.15), we see that the dynamics of h are given by

$$\dot{h}(t) = Q(t)\, l(t). \qquad (3.3.59)$$

Again, if $Q(t) = R(q(t))$, then (3.3.59) may be written as

$$\dot{h} = q * l * q^*. \qquad (3.3.60)$$

Expanding q as $q = q_0 + \vec{q} = q_0 + q_1 i + q_2 j + q_3 k$, we obtain

$$\dot{h} = (q_0 + \vec{q}) * l * (q_0 - \vec{q}) = q_0^2 l + 2q_0 \vec{q} \times l - \vec{q} \times l \times \vec{q}.$$

and

$$\begin{pmatrix} \dot{h}_1 \\ \dot{h}_2 \\ \dot{h}_3 \end{pmatrix} = \begin{pmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2) \\ 2(q_2q_1+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_0q_1) \\ 2(q_3q_1-q_0q_2) & 2(q_3q_2+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}. \quad (3.3.61)$$

For $q \in S^3_+$, q may be parameterized by $\vec{q}$, and it is thus sufficient to consider the dynamics of $\vec{q}$ which read

$$\dot{\vec{q}} = \frac{1}{2}(\sqrt{1-||\vec{q}||^2}\, r + \vec{q} \times r). \quad (3.3.62)$$

The dynamics of h are then given by

$$\dot{h} = (1-\||\vec{q}||^2)l + 2\sqrt{1-||\vec{q}||^2}\, \vec{q} \times l - \vec{q} \times l \times \vec{q}. \quad (3.3.63)$$

(Alternatively, one can substitute $\sqrt{1-(q_1^2+q_2^2+q_3^2)}$ to q_0 in both (3.3.58) and (3.3.61).)

3.3.5 Control system for the underwater vehicule

Using (3.3.63)-(3.3.62) and Proposition 3.3.2, we obtain finally

$$\begin{cases} h' &= q * l * q^*, \\ q' &= \frac{1}{2} q * r, \\ \begin{pmatrix} l \\ r \end{pmatrix}' &= \mathcal{J}^{-1}(Cw' + F(l,r,w)) \end{cases} \quad (3.3.64)$$

where $(h,q,l,r,w) \in \mathbb{R}^3 \times S^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^m$,

$$F(l,r,w) = -\begin{pmatrix} S(r) & 0 \\ S(l) & S(r) \end{pmatrix} \left(\mathcal{J} \begin{pmatrix} l \\ r \end{pmatrix} - Cw \right) - \sum_{p=1}^{n} w_p \begin{pmatrix} L_p^M l + R_p^M r + W_p^M w \\ L_p^J l + R_p^J r + W_p^J w \end{pmatrix}, \quad (3.3.65)$$

and

$$C = -\begin{pmatrix} C^M \\ C^J \end{pmatrix}. \quad (3.3.66)$$

For $q \in \mathcal{O}$, one can replace the two first equations in (3.3.64) by (3.3.63) and (3.3.62), respectively. This results in the system

$$\begin{cases} h' &= (1-\||\vec{q}||^2)l + 2\sqrt{1-||\vec{q}||^2}\, \vec{q} \times l - \vec{q} \times l \times \vec{q}, \\ q' &= \frac{1}{2}(\sqrt{1-||\vec{q}||^2}\, r + \vec{q} \times r), \\ \begin{pmatrix} l \\ r \end{pmatrix}' &= \mathcal{J}^{-1}(Cw' + F(l,r,w)). \end{cases} \quad (3.3.67)$$

3.4 Control properties of the underwater vehicle

3.4.1 Linearization at the equilibrium

When investigating the local controllability of a nonlinear system around an equilibrium point, it is natural to look first at its linearization at the equilibrium point.

To linearize the system (3.3.64) at the equilibrium point $(h, q, l, r, w) = (0, 1, 0, 0, 0)$, we use the parameterization of $\mathcal{O}$ by $\vec{q}$, and consider instead the system (3.3.67).

The linearization of (3.3.67) around $(h, \vec{q}, l, r, w) = (0, 0, 0, 0, 0)$ reads

$$\left\{ \begin{array}{rcl} h' & = & l, \\ 2\vec{q}\,' & = & r, \\ \begin{pmatrix} l \\ r \end{pmatrix}' & = & \mathcal{J}^{-1} C w'. \end{array} \right. \tag{3.4.1}$$

Proposition 3.4.1. *The linearized system* (3.4.1) *with control* $w' \in \mathbb{R}^m$ *is controllable if, and only if,* $rank(C) = 6$.

Proof. The proof follows at once from Kalman rank condition. □

Remark 3.4.2. *The controllability of the linearized system (3.4.1) requires that the number* m *of control inputs fulfills* $m \geq 6$*. If such a condition is fulfilled, we can easily prove the local controllability of the full system (3.3.67).*

3.4.2 Simplications of the model resulting from symmetries

Now we are concerned with the local controllability of (3.3.67) with less than 6 controls inputs. To derive tractable geometric conditions, we consider rigid bodies with symmetries. Let us introduce the operators $S_i(y) = y - 2y_i e_i$, i.e.

$$\begin{array}{l} S_1(y) = (-y_1, y_2, y_3), \\ S_2(y) = (y_1, -y_2, y_3), \\ S_3(y) = (y_1, y_2, -y_3). \end{array} \tag{3.4.2}$$

Definition 3.4.1. *Let* $i \in \{1, 2, 3\}$*. We say that* Ω *is symmetric with respect to the plane* $\{y_i = 0\}$ *if* $S_i(\Omega) = \Omega$*. Let* $f : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$*. If* $f(S_i(y)) = \varepsilon_f^i f(y)$ *for any* $y \in \Omega$ *and some number* $\varepsilon_f^i \in \{-1, 1\}$*, then* f *is said to be even (resp. odd) with respect to* S_i *if* $\varepsilon_f^i = 1$ *(resp.* $\varepsilon_f^i = -1$ *).*

The following proposition gather several useful properties properties of the symmetries S_i, whose proofs are left to the reader.

Proposition 3.4.3. *Let* $i \in \{1, 2, 3\}$*. Then*

1. *$S_iS_i(a) = a$, $\forall a \in \mathbb{R}^3$;*
2. *$S_i(a) \cdot S_i(b) = a \cdot b$, $\forall a, b \in \mathbb{R}^3$;*
3. *$S_i(a) \times S_i(b) = -S_i(a \times b)$, $\forall a, b \in \mathbb{R}^3$;*
4. *If $S_i(\Omega) = \Omega$, then $\hat{n}(S_i(y)) = S_i(\hat{n}(y))$, $\forall y \in \partial\Omega$;*
5. *If $f(S_i(y)) = \varepsilon f(y)$, with $\varepsilon \in \{\pm 1\}$ then $f(S_i(y))\hat{n}(S_i(y)) = \varepsilon S_i(f(y)\hat{n}(y))$;*
6. *If $S_i(\Omega) = \Omega$, then $S_i(y) \times \hat{n}(S_i(y)) = -S_i(y \times \hat{n}(y))$;*
7. *Assume $S_i(\Omega) = \Omega$, and assume given a function $g : \partial\Omega \to \mathbb{R}$ with $g(S_i(y)) = \varepsilon g(y)$, where $\varepsilon \in \{\pm 1\}$. Then the solution f of the system*

$$\begin{cases} \Delta f = 0, & \text{in } \Omega, \\ \dfrac{\partial f}{\partial \hat{n}} = g & \text{on } \partial\Omega, \\ \nabla f(y) \to 0 & \text{as } |y| \to \infty \end{cases}$$

defined up to an additive constant C, fulfills for a convenient choice of C

$$\begin{aligned} f(S_i(y)) &= \varepsilon f(y), \quad y \in \Omega, \\ \nabla f(S_i(y)) &= \varepsilon S_i(\nabla f(y)) \quad y \in \Omega. \end{aligned}$$

8. *Let f and g be as in (7), and let $h(y) = f(y)\partial_{\hat{n}} g(y)$. Then*

$$h(S_p(y)) = \varepsilon_f^p \varepsilon_g^p h(y), \tag{3.4.3}$$

i.e. $\varepsilon^p_{f\partial_{\hat{n}} g} = \varepsilon_f^p \varepsilon_g^p$.

9. *Let f and g be as in (7), and let $h_i(y) = \partial_i f(y)\partial_{\hat{n}} g(y)$. Then*

$$h_i(S_p(y)) = (-1)^{\delta_{ip}} \varepsilon_f^p \varepsilon_g^p h(y), \tag{3.4.4}$$

i.e. $\varepsilon^p_{\partial_i f\partial_{\hat{n}} g} = (-1)^{\delta_{ip}} \varepsilon_f^p \varepsilon_g^p$.

10. *Let f and g be as in (7), and let $h_i(y) = (y \times \nabla f(y))_i \partial_{\hat{n}} g(y)$. Then*

$$h_i(S_p(y)) = -(-1)^{\delta_{ip}} \varepsilon_f^p \varepsilon_g^p h(y), \tag{3.4.5}$$

i.e. $\varepsilon^p_{(y\times\nabla f)_i\partial_{\hat{n}} g} = -(-1)^{\delta_{ip}} \varepsilon_f^p \varepsilon_g^p$.

Applying Proposition 3.4.3 to the solutions ϕ_i, φ_i, $i = 1, 2, 3$, of (3.3.20)-(3.3.22), we obtain at once the following result.

Corollary 3.4.4. *Assume that Ω is symmetric with respect to the plane $\{y_p = 0\}$ (i.e. $S_p(\Omega) = \Omega$) for some $p \in \{1, 2, 3\}$. Then for any $j \in \{1, 2, 3\}$*

$$\begin{aligned} \phi_j(S_p(y)) &= \begin{cases} \phi_j(y) & \text{if } j \neq p, \\ -\phi_j(y) & \text{if } j = p, \end{cases} && (3.4.6) \\ &= (-1)^{\delta_{pj}} \phi_j(y), && (3.4.7) \end{aligned}$$

i.e. $\varepsilon^p_{\phi_j} = (-1)^{\delta_{pj}}$, *and*

$$\begin{aligned}\varphi_j(S_p(y)) &= \begin{cases} -\varphi_j(y) & \text{if } j \neq p, \\ \varphi_j(y) & \text{if } j = p, \end{cases} && (3.4.8)\\ &= -(-1)^{\delta_{pj}} \varphi_j(y), && (3.4.9)\end{aligned}$$

i.e. $\varepsilon^p_{\varphi_j} = -(-1)^{\delta_{pj}}$.

The following result shows how to exploit the symmetries of the rigid body and of the control inputs to simplify the matrices in (3.3.23)-(3.3.30)

Proposition 3.4.5. *Assume that* Ω *is symmetric with respect to the plane* $\{y_p = 0\}$ *for some* $p \in \{1, 2, 3\}$. *Then*

1. $M_{ij} = 0$ *if* $\varepsilon^p_{\phi_i}\varepsilon^p_{\phi_j} = -1$, *i.e.*
$$\delta_{ip} + \delta_{jp} \equiv 1 \pmod 2 \; ; \tag{3.4.10}$$
2. $J_{ij} = 0$ *if* $\varepsilon^p_{\varphi_i}\varepsilon^p_{\varphi_j} = -1$, *i.e.*
$$\delta_{ip} + \delta_{jp} \equiv 1 \pmod 2; \tag{3.4.11}$$
3. $D_{ij} = 0$ *if* $\varepsilon^p_{\phi_i}\varepsilon^p_{\varphi_j} = -1$, *i.e.*
$$\delta_{ip} + \delta_{jp} \equiv 0 \pmod 2; \tag{3.4.12}$$
4. $(C^M)_{ij} = 0$ *if* $\varepsilon^p_{\phi_i}\varepsilon^p_{\chi_j} = -1$, *i.e.*
$$(-1)^{\delta_{ip}} = -\varepsilon^p_{\chi_j}; \tag{3.4.13}$$
5. $(C^J)_{ij} = 0$ *if* $\varepsilon^p_{\varphi_i}\varepsilon^p_{\chi_j} = -1$, *i.e.*
$$(-1)^{\delta_{ip}} = \varepsilon^p_{\chi_j}; \tag{3.4.14}$$
6. $(L^M_q)_{ij} = 0$ *if* $(-1)^{\delta_{ip}}\varepsilon^p_{\phi_j}\varepsilon^p_{\chi_q} = -1$, *i.e.*
$$(-1)^{\delta_{ip}+\delta_{jp}} = -\varepsilon^p_{\chi_q}; \tag{3.4.15}$$
7. $(R^M_q)_{ij} = 0$ *if* $(-1)^{\delta_{ip}}\varepsilon^p_{\varphi_j}\varepsilon^p_{\chi_q} = -1$, *i.e*
$$(-1)^{\delta_{ip}+\delta_{jp}} = \varepsilon^p_{\chi_q}; \tag{3.4.16}$$
8. $(W^M_q)_{ij} = 0$ *if* $(-1)^{\delta_{ip}}\varepsilon^p_{\varphi_j}\varepsilon^p_{\chi_q} = -1$, *i.e.*
$$(-1)^{\delta_{ip}} = -\varepsilon^p_{\chi_j}\varepsilon^p_{\chi_q}; \tag{3.4.17}$$

9. $(L_q^J)_{ij} = 0$ *if* $-(-1)^{\delta_{ip}}\varepsilon^p_{\phi_j}\varepsilon^p_{\chi_q} = -1$*, i.e.*

$$(-1)^{\delta_{ip}+\delta_{jp}} = \varepsilon^p_{\chi_q}; \tag{3.4.18}$$

10. $(R_q^J)_{ij} = 0$ *if* $-(-1)^{\delta_{ip}}\varepsilon^p_{\varphi_j}\varepsilon^p_{\chi_q} = -1$*, i.e.*

$$(-1)^{\delta_{ip}+\delta_{jp}} = -\varepsilon^p_{\chi_q}; \tag{3.4.19}$$

11. $(W_q^J)_{ij} = 0$ *if*

$$(-1)^{\delta_{ip}} = \varepsilon^p_{\chi_j}\varepsilon^p_{\chi_q}, \tag{3.4.20}$$

where the matrices $M, J^R, D, C^M, C^J, L_q^M, R_q^M, L_q^J, R_q^J, W_q^M$ *and* W_q^J *are defined in* (3.3.23)-(3.3.30).

From now on, we assume that Ω is invariant under the operators S_2 and S_3, i.e.

$$S_p(\Omega) = \Omega, \quad \forall p \in \{2,3\}, \tag{3.4.21}$$

and that $\varepsilon^p_{\chi_1} = 1$, i.e.

$$\chi_1(S_p(y)) = \chi_1(y) \quad \forall y \in \partial\Omega, \forall p \in \{2,3\}. \tag{3.4.22}$$

In other words, the set S and the control χ_1 are symmetric with respect to the two planes $\{y_2 = 0\}$ and $\{y_3 = 0\}$. As a consequence, several coefficients in the matrices in (3.3.23)-(3.3.30) vanish.

More precisely, using (3.4.21)-(3.4.22) and Proposition 3.4.5, we see immediately that the matrices in (3.3.33) can be written

$$M = \begin{pmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & 0 \\ 0 & 0 & M_{33} \end{pmatrix}, \quad J = \begin{pmatrix} J_{11} & 0 & 0 \\ 0 & J_{22} & 0 \\ 0 & 0 & J_{33} \end{pmatrix}, \tag{3.4.23}$$

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & D_{23} \\ 0 & D_{32} & 0 \end{pmatrix}, \tag{3.4.24}$$

$$C^M e_1 = \begin{pmatrix} (C^M)_{11} \\ 0 \\ 0 \end{pmatrix}, \quad C^J e_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{3.4.25}$$

$$L_1^M = \begin{pmatrix} (L_1^M)_{11} & 0 & 0 \\ 0 & (L_1^M)_{22} & 0 \\ 0 & 0 & (L_1^M)_{33} \end{pmatrix}, \quad R_1^M = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & (R_1^M)_{23} \\ 0 & (R_1^M)_{32} & 0 \end{pmatrix} \tag{3.4.26}$$

$$(W_1^M)e_1 = \begin{pmatrix} (W_1^M)_{11} \\ 0 \\ 0 \end{pmatrix}, \quad L_1^J = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & (L_1^J)_{23} \\ 0 & (L_1^J)_{32} & 0 \end{pmatrix}, \tag{3.4.27}$$

and

$$R_1^J = \begin{pmatrix} (R_1^J)_{11} & 0 & 0 \\ 0 & (R_1^J)_{22} & 0 \\ 0 & 0 & (R_1^J)_{33} \end{pmatrix}, \quad (W_1^J)e_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.4.28}$$

3.4.3 Toy problem

Before investigating the full system (3.3.67), it is very important to look at the simplest situation for which $h_i = r_i = 0$ for $i = 2, 3$, $\vec{q} = 0$, $r = 0$, and $w_j = 0$ for $j = 2, .., m$.

Lemma 3.4.6. *Assume that* (3.4.21)-(3.4.22) *hold, and assume given some functions* $h_1, l_1, w_1 \in C^1([0,T])$ *satisfying*

$$\begin{cases} h_1' &= l_1 \\ \\ l_1' &= \alpha w_1' + \beta l_1 w_1 + \gamma (w_1)^2, \end{cases} \tag{3.4.29}$$

where

$$\alpha := \frac{-(C^M)_{11}}{m + M_{11}}, \quad \beta := \frac{-(L_1^M)_{11}}{m + M_{11}}, \quad \textit{and } \gamma := \frac{-(W_1^M)_{11}}{m + M_{11}}.$$

Let $h := (h_1, 0, 0)$, $\vec{q} := (0, 0, 0)$, $l := (l_1, 0, 0)$, $r := (0, 0, 0)$, *and* $w := (w_1, 0, ..., 0)$. *Then* $(h, \vec{q}, l, r, w)$ *solves* (3.3.67).

Proof. Let us set $h = h_1 e_1$, $\vec{q} = 0$, $l = l_1 e_1$, $r = 0$ and $w = (w_1, 0, ..., 0)$, where (h_1, l_1, w_1) fulfills (3.4.29). From (3.4.23)-(3.4.25), we have that

$$\mathcal{J}\begin{pmatrix} l \\ r \end{pmatrix} = l_1 \mathcal{J} e_1 = l_1 \begin{pmatrix} m + M_{11} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{3.4.30}$$

and

$$Cw = w_1 C e_1 = -w_1 \begin{pmatrix} (C^M)_{11} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.4.31}$$

This yields

$$\begin{pmatrix} S(r) & 0 \\ \\ S(l) & S(r) \end{pmatrix} \left(\mathcal{J} \begin{pmatrix} l \\ r \end{pmatrix} - Cw \right) = 0. \tag{3.4.32}$$

Replacing in (3.3.65), we obtain

$$
\begin{aligned}
F(l,r,w) &= -\sum_{p=1}^{n} w_p \begin{pmatrix} L_p^M l + R_p^M r + W_p^M w \\ \\ L_p^J l + R_p^J r + W_p^J w \end{pmatrix} \\
&= -w_1 \begin{pmatrix} L_1^M l + W_1^M w \\ \\ L_1^J l + W_1^J w \end{pmatrix} \\
&= -w_1 \left(l_1 \begin{pmatrix} (L_1^M)_{11} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + w_1 \begin{pmatrix} (W_1^M)_{11} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right),
\end{aligned}
\tag{3.4.33}
$$

We conclude that $(h, \vec{q}, l, r, w)$ is a solution of (3.3.67). □

3.4.4 Return method

The main result in this section (see below Theorem 3.4.9) is derived in following a strategy developed in [45] and inspired in part from Coron's return method. We first construct a (non trivial) loop-shaped trajectory of the control system (3.3.67), which is based on the computations performed in Lemma 3.4.6. Next, we compute the linearized system along the above reference trajectory. Finally, we use a controllability test from [45] to investigate the controllability of the linearized system, in which the control appears with its time derivative.

Construction of a loop-shaped trajectory.

The construction differs slightly from those in [45]: indeed, to simplify the computations, we impose here that all the derivatives of $\bar{l}_1$ of order larger than two vanish at $t = T$. For given $T > 0$, we pick any function $\xi \in C^\infty(\mathbb{R}; [0,1]$ such that

$$
\xi(t) = \begin{cases} 0 & \text{if } \ t < \dfrac{T}{3}, \\[2mm] 1 & \text{if } \ t > \dfrac{2T}{3}. \end{cases}
$$

Pick any $\lambda_0 > 0$ and let $\lambda \in [-\lambda_0, \lambda_0]$ with $\lambda \neq 0$. Set

$$
\overline{h}_1(t) = \lambda \xi(t)(t-T)^2, \quad l_1(t) = \overline{h}_1'(t), \qquad t \in \mathbb{R}. \tag{3.4.34}
$$

Note that

$$
\overline{h}_1(0) = \overline{h}_1(T) = \bar{l}_1(0) = \bar{l}_1(T) = 0 \tag{3.4.35}
$$

$$
\bar{l}_1'(T) = 2\lambda \neq 0, \quad \bar{l}_1^{(k)}(T) = 0 \ \text{ for } k \geq 2, \tag{3.4.36}
$$

and hence

$$\hat{A}^{(k)}(T) = \hat{B}^{(k)}(T) = \hat{D}^{(k)}(T) = 0 \quad \text{for } k \in \mathbb{N} \setminus \{0\}. \tag{3.4.37}$$

Next, define $\overline{w}_1$ as the solution to the Cauchy problem

$$\begin{aligned} \dot{\overline{w}}_1 &= \alpha^{-1}(\dot{\bar{l}}_1 - \beta \bar{l}_1 \overline{w}_1 - \gamma \overline{w}_1^2) & (3.4.38)\\ \overline{w}_1(0) &= 0. & (3.4.39) \end{aligned}$$

By a classical result on the continuous dependence of solutions of ODE with respect to a parameter, we have that the solution $\overline{w}_1$ of (3.4.38)-(3.4.39) is defined on $[0, T]$ provided that λ_0 is small enough. Set $\overline{h} = (h_1, 0, 0)$, $\overline{\vec{q}} = (0, 0, 0)$, $\overline{w} = (w_1, 0, ..., 0)$, $\bar{l} = (l_1, 0, 0)$ and $\overline{r} = (0, 0, 0)$. According to Lemma 3.4.6, $(\overline{h}, \overline{\vec{q}}, \bar{l}, \overline{r}, \overline{w})$ is a solution of (3.3.67), which satisfies

$$(\overline{h}, \overline{\vec{q}}, \bar{l}, \overline{r})(0) = 0 = (\overline{h}, \overline{\vec{q}}, \bar{l}, \overline{r})(T).$$

Linearization along the reference trajectory

Writing

$$\begin{aligned} h &= \overline{h} + \hat{h},\\ \vec{q} &= \overline{\vec{q}} + \widehat{\vec{q}},\\ l &= \bar{l} + \hat{l},\\ r &= \overline{r} + \hat{r}, \end{aligned} \tag{3.4.40}$$

expanding in (3.3.67) in keeping only the first order terms in $\hat{h}, \widehat{\vec{q}}, \hat{l}$ and $\hat{r}$, we obtain the following linear system

$$\begin{cases} \hat{h}' &= \hat{l} + 2\widehat{\vec{q}} \times \bar{l},\\ \widehat{\vec{q}}\,' &= \frac{1}{2}\hat{r},\\ \begin{pmatrix} \hat{l} \\ \hat{r} \end{pmatrix}' &= \mathcal{J}^{-1}\left(A(t)\begin{pmatrix} \hat{l} \\ \hat{r} \end{pmatrix} + B(t)\hat{w} + C\hat{w}'\right), \end{cases} \tag{3.4.41}$$

where the matrices $A(t) \in \mathbb{R}^{6\times 6}$ and $B(t) \in \mathbb{R}^{6\times m}$ are defined as

$$\begin{aligned} A(t) &= \left(\frac{\partial F}{\partial l}(\bar{l}(t), \overline{r}(t), \overline{w}(t)) \quad \Big| \quad \frac{\partial F}{\partial r}(\bar{l}(t), \overline{r}(t), \overline{w}(t))\right), & (3.4.42)\\ B(t) &= \frac{\partial F}{\partial w}(\bar{l}(t), \overline{r}(t), \overline{w}(t)). & (3.4.43) \end{aligned}$$

Setting

$$\hat{p} = 2\widehat{\vec{q}}, \tag{3.4.44}$$

we can write (3.4.41) as

$$\begin{cases} \hat{h}' &= \hat{l} - \bar{l}_1 e_1 \times \hat{p},\\ \hat{p}' &= \hat{r},\\ \begin{pmatrix} \hat{l} \\ \hat{r} \end{pmatrix}' &= \mathcal{J}^{-1}\left(A(t)\begin{pmatrix} \hat{l} \\ \hat{r} \end{pmatrix} + B(t)\hat{w} + C\hat{w}'\right), \end{cases} \tag{3.4.45}$$

Obviously, (3.4.41) is controllable on $[0,T]$ if, and only if, (3.4.45) is. Letting

$$z = \begin{pmatrix} \hat{h} \\ \hat{p} \end{pmatrix}, \quad k = \begin{pmatrix} \hat{l} \\ \hat{r} \end{pmatrix}, \quad f = \hat{w},$$

we obtain the following control system

$$\begin{aligned} \begin{pmatrix} \dot{z} \\ \dot{k} \end{pmatrix} &= \begin{pmatrix} D(t) & Id \\ 0 & \mathcal{J}^{-1}A(t) \end{pmatrix} \begin{pmatrix} z \\ k \end{pmatrix} + \begin{pmatrix} 0 \\ \mathcal{J}^{-1}B(t) \end{pmatrix} f + \begin{pmatrix} 0 \\ \mathcal{J}^{-1}C \end{pmatrix} \dot{f} \\ &=: \mathcal{A}(t) \begin{pmatrix} z \\ k \end{pmatrix} + \mathcal{B}(t) f + \mathcal{C}\dot{f}. \end{aligned} \tag{3.4.46}$$

We find that

$$D = \begin{pmatrix} 0 & -S(\bar{l}) \\ 0 & 0 \end{pmatrix}, \qquad \text{with } S(\bar{l}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\bar{l}_1 \\ 0 & \bar{l}_1 & 0 \end{pmatrix},$$

$$Bf = \begin{pmatrix} 0 & 0 \\ S(\bar{l}) & 0 \end{pmatrix} Cf - \overline{w}_1 \begin{pmatrix} W_1^M \\ W_1^J \end{pmatrix} f - \sum_{p=1}^{m} f_p \begin{pmatrix} L_p^M & W_p^M \\ L_p^J & W_p^J \end{pmatrix} \begin{pmatrix} \bar{l} \\ \overline{w} \end{pmatrix},$$

and that

$$A = \begin{pmatrix} -(L_1^M)_{11}\overline{w}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -(L_1^M)_{22}\overline{w}_1 & 0 & 0 & 0 & A_{26} \\ 0 & 0 & -(L_1^M)_{33}\overline{w}_1 & 0 & A_{35} & 0 \\ 0 & 0 & 0 & -(R_1^J)_{11}\overline{w}_1 & 0 & 0 \\ 0 & 0 & A_{53} & 0 & D_{32}\bar{l}_1 - (R_1^J)_{22}\overline{w}_1 & 0 \\ 0 & A_{62} & 0 & 0 & 0 & -D_{32}\bar{l}_1 - (R_1^J)_{33}\overline{w}_1 \end{pmatrix}$$

with

$$\begin{aligned} A_{26} &= -(m + M_{11})\overline{l_1} - \big((C^M)_{11} + (R_1^M)_{23}\big)\overline{w}_1, \\ A_{35} &= (m + M_{11})\overline{l_1} + \big((C^M)_{11} - (R_1^M)_{32}\big)\overline{w}_1, \\ A_{53} &= (M_{33} - M_{11})\overline{l_1} - \big((C^M)_{11} + (L_1^J)_{33}\big)\overline{w}_1, \\ A_{62} &= (M_{11} - M_{22})\overline{l_1} + \big((C^M)_{11} - (L_1^J)_{32}\big)\overline{w}_1. \end{aligned}$$

From now on, we suppose in addition to (3.4.21)-(3.4.22) that χ_1 is chosen so that

$$\alpha \neq 0. \tag{3.4.47}$$

Linear control systems with one derivative in the control

Let us consider any linear control system of the form

$$\dot{x} = \mathcal{A}(t)x + \mathcal{B}(t)u + C\dot{u} \tag{3.4.48}$$

where $x \in \mathbb{R}^n$ is the state ($n \geq 1$), $u \in \mathbb{R}^m$ is the control input ($m \geq 1$), $\mathcal{A} \in C^\infty([0,T];\mathbb{R}^{n\times n})$, $\mathcal{B} \in C^\infty([0,T];\mathbb{R}^{n\times m})$, and $\mathcal{C} \in \mathbb{R}^{n\times m}$. Define a sequence of matrices $\mathcal{M}_i(t) \in \mathbb{R}^{n\times m}$ by

$$\mathcal{M}_0(t) = \mathcal{B}(t) + \mathcal{A}(t)\mathcal{C}, \quad \text{and} \quad \mathcal{M}_i(t) = \dot{\mathcal{M}}_{i-1}(t) - \mathcal{A}\mathcal{M}_{i-1}(t), \qquad \forall i \geq 1,\ \forall t \geq 0. \tag{3.4.49}$$

Introduce the reachable set

$$\mathcal{R}_{u(0)=0} = \{x_T \in \mathbb{R}^n; \exists u \in H^1(0,T;\mathbb{R}^m) \text{ with } u(0)=0 \text{ such that } x_T = x(T), \text{ where } x(\cdot) \text{ solves (3.4.48) and } x(0)=0\}.$$

Then the following result holds.

Proposition 3.4.7. *[45, Proposition 2.12] Let $\varepsilon > 0$, $\mathcal{A} \in C^\omega((-\varepsilon, T+\varepsilon);\mathbb{R}^{n\times n})$ and $\mathcal{B} \in C^\omega((-\varepsilon, T+\varepsilon);\mathbb{R}^{n\times m})$, and let $(\mathcal{M}_i)_{i\geq 0}$ be the sequence defined in (3.4.49). Then for all $t_0 \in [0,T]$, we have that*

$$\mathcal{R}_{u(0)=0} = \mathcal{C}\mathbb{R}^m + Span\{\phi(T,t_0)\,\mathcal{M}_i(t_0)u;\ u \in \mathbb{R}^m, i \geq 0\}, \tag{3.4.50}$$

where ϕ denotes the fundamental solution associated with the system $\dot{x} = \mathcal{A}(t)x$.

Recall that the fundamental solution associated with $\dot{x} = \mathcal{A}(t)x$ is defined as the solution to

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \mathcal{A}(t)\phi(t,s), \\ \phi(s,s) &= Id. \end{aligned}$$

For notational convenience, we introduce the matrices

$$\hat{A}(t) = \mathcal{J}^{-1}A(t), \quad \hat{B}(t) = \mathcal{J}^{-1}B(t), \quad \hat{C} = \mathcal{J}^{-1}C, \quad \mathcal{M}_i(t) = \begin{pmatrix} U_i(t) \\ V_i(t) \end{pmatrix}, \tag{3.4.51}$$

where $\hat{A}(t) \in \mathbb{R}^{6\times 6}$, $\hat{B}(t), \hat{C}, U_i(t), V_i(t) \in \mathbb{R}^{6\times m}$. Then

$$\begin{pmatrix} U_0(t) \\ V_0(t) \end{pmatrix} = \begin{pmatrix} \hat{C} \\ \hat{B}(t) + \hat{A}(t)\hat{C} \end{pmatrix}, \tag{3.4.52}$$

while

$$\begin{pmatrix} U_i(t) \\ V_i(t) \end{pmatrix} = \begin{pmatrix} U'_{i-1}(t) - D(t)U_{i-1}(t) - V_{i-1}(t) \\ V'_{i-1}(t) - \hat{A}(t)V_{i-1}(t) \end{pmatrix}. \tag{3.4.53}$$

The following result reveals that half of the terms $U_i(t)$ and $V_i(t)$ vanish at $t = T$.

Proposition 3.4.8. *If $\hat{C} \in \mathbb{R}^{6\times m}$ is given and $\hat{A}, \hat{B}, \hat{D}$ denote functions in $C^\infty([0,T];\mathbb{R}^{6\times 6})$ fulfilling*

$$\hat{A}^{(2l)}(T) = \hat{B}^{(2l)}(T) = \hat{D}^{(2l)}(T) = 0 \qquad \forall l \in \mathbb{N} \tag{3.4.54}$$

then the sequences $(U_i)_{i\geq 0}$ and $(V_i)_{i\geq 0}$ defined in (3.4.52)-(3.4.53) satisfy

$$V_{2k}^{(2l)}(T) = V_{2k+1}^{(2l+1)}(T) = 0, \qquad \forall k,l \in \mathbb{N} \tag{3.4.55}$$

$$U_{2k+1}^{(2l)}(T) = U_{2k}^{(2l+1)}(T) = 0, \qquad \forall k,l \in \mathbb{N}. \tag{3.4.56}$$

The following result, which is the main result in the paper, shows that under suitable assumptions the local controllability of (3.3.67) holds with less than six control inputs.

Theorem 3.4.9. *Assume that* (3.4.21), (3.4.22) *and* (3.4.47) *hold. If both rank conditions*

$$\text{rank } (\hat{C}, V_1(T), V_3(T), V_5(T), ...) = 6, \tag{3.4.57}$$

and

$$\text{rank } (\hat{C}, U_2(T), U_4(T), U_6(T), ...) = 6, \tag{3.4.58}$$

are fulfilled, then for any $T > 0$ the system (3.3.67), with state $(h, \vec{q}, l, r) \in \mathbb{R}^{12}$ and control $w \in \mathbb{R}^m$, is locally controllable around the origin in time T. We can also impose that the control input $w \in H^1(0, T; \mathbb{R}^m)$ satisfies $w(0) = 0$. Moreover, for some $\eta > 0$, there is a C^1 map from $B_{\mathbb{R}^{24}}(0, \eta)$ to $H^2(0, T; \mathbb{R}^m)$, which associates to $(h_0, \vec{q}_0, l_0, r_0, h_T, \vec{q}_T, l_T, r_T)$ a control satisfying $w(0) = 0$ and steering the state of the system from $(h_0, \vec{q}_0, l_0, r_0)$ at $t = 0$ to $(h_T, \vec{q}_T, l_T, r_T)$ at $t = T$.

Proof. **Step 1: Controllability of the linearized system.**
Letting $t_0 = T$ in Proposition 3.4.7 yields

$$\begin{aligned}
\mathcal{R}_{f(0)=0} &= \mathcal{C}\mathbb{R}^m + \sum_{i\geq 0} \mathcal{M}_i(T)\mathbb{R}^m \\
&= \begin{pmatrix} 0 \\ \hat{C} \end{pmatrix} \mathbb{R}^m + \sum_{j\geq 0} \begin{pmatrix} 0 \\ V_{2j+1}(T) \end{pmatrix} \mathbb{R}^m + \begin{pmatrix} \hat{C} \\ 0 \end{pmatrix} \mathbb{R}^m + \sum_{j\geq 0} \begin{pmatrix} U_{2j}(T) \\ 0 \end{pmatrix} \mathbb{R}^m.
\end{aligned}$$

Thus, if the conditions (3.4.57) and (3.4.58) are fulfilled, we infer that $\mathcal{R}_{f(0)=0} = \mathbb{R}^{12}$, i.e. the system (3.4.45) is controllable. The same is true for (3.4.41).

Step 2: Local controllability of the nonlinear system.
Let us introduce the Hilbert space

$$\mathcal{H} := \mathbb{R}^{12} \times \{f \in H^2(0, T; \mathbb{R}^m);\ f(0) = 0\}$$

endowed with its natural Hilbertian norm

$$\|(x, f)\|^2_{\mathcal{H}} = |x|^2_{\mathbb{R}^{12}} + \|f\|^2_{H^2(0,T)}\,.$$

We denote by $B_{\mathcal{H}}(0, \delta)$ the open ball in $\mathcal{H}$ with center 0 and radius δ, i.e.

$$B_{\mathcal{H}}(0, \delta) = \{(x, f) \in \mathcal{H}; \|(x, f)\|_{\mathcal{H}} < \delta\}.$$

Let us introduce the map

$$\begin{aligned}
\Gamma : B_{\mathcal{H}}(0, \delta) &\to \mathbb{R}^{24} \\
((h_0, \vec{q}_0, l_0, r_0), f) &\mapsto (h_0, \vec{q}_0, l_0, r_0, h(T), \vec{q}(T), l(T), r(T))
\end{aligned}$$

where $(h(t), \vec{q}(t), l(t), r(t))$ denotes the solution of

$$\left\{ \begin{aligned}
h' &= (1 - \||\vec{q}\|^2)l + 2\sqrt{1 - ||\vec{q}||^2}\, \vec{q} \times l - \vec{q} \times l \times \vec{q}, \\
q' &= \tfrac{1}{2}(\sqrt{1 - ||\vec{q}||^2}\, r + \vec{q} \times r), \\
\begin{pmatrix} l \\ r \end{pmatrix}' &= \mathcal{J}^{-1}(C(\overline{w}' + f') + F(l, r, \overline{w} + f)), \\
(h(0), \vec{q}(0), l(0), r(0)) &= (h_0, \vec{q}_0, l_0, r_0).
\end{aligned} \right. \tag{3.4.59}$$

Note that Γ is well defined for $\delta > 0$ small enough. Using the Sobolev embedding $H^2(0,T) \subset C^1([0,T])$, we can prove as in [87, Theorem 1] that Γ is of class C^1 on $B_{\mathcal{H}}(0,\delta)$ and that its tangent linear map at the origin is given by

$$d\Gamma(0)((\hat{h}_0, \widehat{\vec{q}_0}, \hat{l}_0, \hat{r}_0), f) = (\hat{h}_0, \widehat{\vec{q}_0}, \hat{l}_0, \hat{r}_0, \hat{h}(T), \widehat{\vec{q}}(T), \hat{l}(T), \hat{r}(T)),$$

where $(\hat{h}(t), \widehat{\vec{q}}(t), \hat{l}(t), \hat{r}(t))$ solves the system (3.4.41) with initial conditions

$$(\hat{h}(0), \widehat{\vec{q}}(0), \hat{l}(0), \hat{r}(0)) = (\hat{h}_0, \widehat{\vec{q}_0}, \hat{l}_0, \hat{r}_0).$$

We know from Step 2 that (3.4.41) is controllable, so that $d\Gamma(0)$ is onto. Let $V := (\ker d\Gamma(0))^{\perp}$ denote the orthogonal complement of $\ker d\Gamma(0)$ in $\mathcal{H}$. Then $d\Gamma|_V(0)$ is invertible, and therefore it follows from the inverse function theorem that the map $\Gamma|_V : V \to \mathbb{R}^{24}$ is locally invertible at the origin. More precisely, there exists a number $\delta > 0$ and an open set $A \subset \mathbb{R}^{24}$ containing 0, such that the map $\Gamma : B_{\mathcal{H}}(0,\delta) \cap V \to \Omega$ is well-defined, of class C^1, invertible, and with an inverse map of class C^1. Let us denote this inverse map by Γ^{-1}, and let us write $\Gamma^{-1}(x_0, x_T) = (x_0, f(x_0, x_T))$. Finally, let us set $w = \overline{w} + f$. Then, for $\eta > 0$ small enough, we have that

$$w \in C^1(B_{\mathbb{R}^{24}}(0,\eta), H^2(0,T,\mathbb{R}^m)), \tag{3.4.60}$$

and that the solution $(h(t), \vec{q}(t), l(t), r(t))$ of system (3.3.67), with the initial conditions

$$(h(0), \vec{q}(0), l(0), r(0)) = (h_0, \vec{q}_0, l_0, r_0),$$

satisfies

$$(h(T), \vec{q}(T), l(T), r(T)) = (h_T, \vec{q}_T, l_T, r_T).$$

The proof of Theorem 3.4.9 is complete. $\square$

We now derive two corollaries of Theorem 3.4.9, that will be used in the next section. We introduce the matrices

$$\mathbf{A} = \Big(A_L \Big| A_R \Big), \tag{3.4.61}$$

where

$$A_L = \begin{pmatrix} -(L_1^M)_{11} & 0 & 0 \\ 0 & -(L_1^M)_{22} & 0 \\ 0 & 0 & -(L_1^M)_{33} \\ 0 & 0 & 0 \\ 0 & 0 & \alpha(M_{33} - M_{11}) - ((L_1^J)_{23} + (C^M)_{11}) \\ 0 & \alpha(M_{11} - M_{22}) - ((L_1^J)_{32} - (C^M)_{11}) & 0 \end{pmatrix}, \tag{3.4.62}$$

$$A_R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\alpha\mathcal{J}_{11} - ((R_1^M)_{23} + (C^M)_{11}) \\ 0 & \alpha\mathcal{J}_{11} - ((R_1^M)_{32} - (C^M)_{11}) & 0 \\ -(R_1^J)_{11} & 0 & 0 \\ 0 & \alpha D_{32} - (R_1^J)_{22} & 0 \\ 0 & 0 & -\alpha D_{23} - (R_1^J)_{33} \end{pmatrix}, \tag{3.4.63}$$

$$\begin{aligned}\mathbf{B} \;=\; & \begin{pmatrix} 0 \\ \\ -\alpha S(e_1)C^M \end{pmatrix} - \alpha \left(\begin{array}{c|c|c|c} L_1^M e_1 & L_2^M e_1 & \cdots & L_n^M e_1 \\ & & & \\ L_1^J e_1 & L_2^J e_1 & \cdots & L_n^J e_1 \end{array}\right) - \begin{pmatrix} W_1^M \\ \\ W_1^J \end{pmatrix} \\ & - \left(\begin{array}{c|c|c|c} W_1^M e_1 & W_2^M e_1 & \cdots & W_n^M e_1 \\ & & & \\ W_1^J e_1 & W_2^J e_1 & \cdots & W_n^J e_1 \end{array}\right), \end{aligned} \tag{3.4.64}$$

and

$$\mathbf{D} = \begin{pmatrix} 0 & -\alpha S(e_1) \\ 0 & 0 \end{pmatrix}. \tag{3.4.65}$$

The first corollary will give a controllability result with four controls inputs.

Corollary 3.4.10. *If the conditions*

$$\text{rank } (C, \mathbf{B} + \mathbf{A}\mathcal{J}^{-1}C) = 6, \tag{3.4.66}$$

and

$$\text{rank } (C, \frac{1}{2}\mathcal{J}\mathbf{D}\mathcal{J}^{-1}C + \mathbf{B} + \mathbf{A}\mathcal{J}^{-1}C) = 6, \tag{3.4.67}$$

are fulfilled, then the conditions (3.4.57)-(3.4.58) *are satisfied, so that the conclusion of Theorem 3.4.9 is valid.*

Proof. We distinguish two cases.
Case 1: $\gamma + \alpha\beta = 0$.
We begin with the "simplest" case when $\gamma + \alpha\beta = 0$. Let $\overline{g}_1 := \bar{l}_1 - \alpha\overline{w}_1$. It is clear that $\dot{\overline{g}}_1 = \beta\overline{w}_1\overline{g}_1$, hence $\overline{g}_1 \equiv 0$. We infer that

$$\begin{aligned}&\overline{w}_1^{(k)}(T) = \alpha^{-1}\bar{l}_1^{(k)}(T) = 0 \quad \text{ for } \; k \in \mathbb{N}\setminus\{1\}, \\ &\overline{w}_1'(T) = \alpha^{-1}\bar{l}_1'(T) = 2\lambda/\alpha \neq 0.\end{aligned}$$

It follows that

$$\begin{aligned}&A^{(k)}(T) = 0, \quad B^{(k)}(T) = 0, \quad D^{(k)}(T) = 0 \qquad \text{for } \; k \in \mathbb{N}\setminus\{1\}, \\ &A'(T) = \overline{w}_1'(T)\mathbf{A}, \quad B'(T) = \overline{w}_1'(T)\mathbf{B}, \quad D'(T) = \overline{w}_1'(T)\mathbf{D}.\end{aligned}$$

On the other hand, it is easily seen that

$$\begin{aligned}&V_1(T) = V_0'(T) = \mathcal{J}^{-1}B'(T) + \mathcal{J}^{-1}A'(T)\mathcal{J}^{-1}C = \overline{w}_1'(T)\left(\mathcal{J}^{-1}\mathbf{B} + \mathcal{J}^{-1}\mathbf{A}\mathcal{J}^{-1}C\right) \\ &U_2(T) = -D'(T)U_0(T) - 2V_0'(T) = -\overline{w}_1'(T)[\mathbf{D}\mathcal{J}^{-1}C + 2\mathcal{J}^{-1}(\mathbf{B} + \mathbf{A}\mathcal{J}^{-1}C)].\end{aligned}$$

Therefore

$$\begin{aligned}&\text{rank}(\hat{C}, V_1(T)) = \text{rank } (C, \mathbf{B} + \mathbf{A}\mathcal{J}^{-1}C) = 6, \\ &\text{rank}(\hat{C}, U_2(T)) = \text{rank } (C, \frac{1}{2}\mathcal{J}\mathbf{D}\mathcal{J}^{-1}C + \mathbf{B} + \mathbf{A}\mathcal{J}^{-1}C) = 6,\end{aligned}$$

as desired.

Case 2. $\gamma+\alpha\beta\neq 0$. We claim that for $T>0$ arbitrary chosen and λ_0 small enough, we have for $0<\lambda<\lambda_0$,

$$\text{rank }(\mathcal{C},\mathcal{M}_0(T),\mathcal{M}_1(T),\mathcal{M}_2(T))=12.$$

First, $\|\overline{p}_1\|_{W^{2,\infty}(0,T)}=O(\lambda)$ still with $\overline{p}_1(T)=\ddot{\overline{p}}_1(T)=0$. From (3.4.38)-(3.4.39), we infer with Gronwall lemma (for λ_0 small enough) that $\overline{w}_1$ is well defined on $[0,T]$ and that $\|\overline{w}_1\|_{L^\infty(0,T)}=O(\lambda)$. This also yields (with (3.4.38)) $||\overline{w}_1||_{W^{2,\infty}(0,T)}=O(\lambda)$. Next, integrating in (3.4.38) over $(0,T)$ yields $\overline{w}_1(T)=O(\lambda^2)$. Finally, derivating in (3.4.38) gives $\ddot{\overline{w}}_1(T)=O(\lambda^2)$. We conclude that

$$(A(T),B(T),\ddot{A}(T),\ddot{B}(T))=O(\lambda^2),\quad D(T)=0,$$

while

$$(\dot{A}(T),\dot{B}(T),\dot{D}(T))=\alpha^{-1}(2\pi/T)^2\lambda(\mathbf{A},\mathbf{B},\mathbf{D})+O(\lambda^4),$$

for $\dot{\overline{p}}_1(T)=\alpha\dot{\overline{w}}_1(T)+O(\lambda^4)$. It follows that

$$\begin{aligned}
&\text{rank}\,(\mathcal{C},\mathcal{M}_0(T),\mathcal{M}_1(T),\mathcal{M}_2(T))\\
&\quad=\text{rank}\left[\begin{pmatrix}0\\ \mathcal{J}^{-1}C\end{pmatrix},\begin{pmatrix}\mathcal{J}^{-1}C\\ 0\end{pmatrix},\right.\\
&\qquad\left.\begin{pmatrix}0\\ \mathcal{J}^{-1}(\mathbf{B}+\mathbf{A}\mathcal{J}^{-1}C)\end{pmatrix},\begin{pmatrix}\mathcal{J}^{-1}[\mathcal{J}\mathbf{D}\mathcal{J}^{-1}C+2(\mathbf{B}+\mathbf{A}\mathcal{J}^{-1}C)]\\ 0\end{pmatrix}\right]\\
&\quad=12,
\end{aligned}$$

for $0<\lambda<\lambda_0$ with λ_0 small enough. This proves that (3.3.67) is controllable. □

The second one is based on the explicit computations of $\mathcal{M}_i(T)$ for $i\leq 8$. It will be used later to derive a controllability result with only three controls inputs.

Corollary 3.4.11. *Let* $\mathbf{E}:=\mathbf{B}+\mathbf{A}\mathcal{J}^{-1}C$. *If the conditions*

$$\textit{rank }(C,\mathbf{E},\mathbf{A}\mathcal{I}^{-1}\mathbf{E},(\mathbf{A}\mathcal{I}^{-1})^2\mathbf{E},(\mathbf{A}\mathcal{J}^{-1})^3\mathbf{E})=6,\tag{3.4.68}$$

and

$$\begin{aligned}
&\textit{rank }(C,\frac{1}{2}\mathcal{J}\mathbf{D}\mathcal{J}^{-1}C+\mathbf{E},(\mathcal{J}\mathbf{D}\mathcal{J}^{-1}+2\mathbf{A}\mathcal{J}^{-1})\mathbf{E},\\
&\qquad(2\mathcal{J}\mathbf{D}\mathcal{J}^{-1}+3\mathbf{A}\mathcal{J}^{-1})\mathbf{A}\mathcal{J}^{-1}\mathbf{E},(15\mathcal{J}\mathbf{D}\mathcal{J}^{-1}+48\mathbf{A}\mathcal{J}^{-1})(\mathbf{A}\mathcal{J}^{-1})^2\mathbf{E},)=6,
\end{aligned}\tag{3.4.69}$$

are fulfilled, then the conditions (3.4.57)-(3.4.58) *are satisfied, so that the conclusion of Theorem 3.4.9 is valid.*

Proof. The proof is almost the same as for Corollary 3.4.10, the only difference being that we need now to compute $\mathcal{M}_i(T)$ for $i\leq 8$. In view of Proposition 3.4.8, it is sufficient to compute $V_i(T)$ for $i\in\{1,3,5,7\}$ and $U_i(T)$ for $i\in\{2,4,6,8\}$. The results are displayed in two propositions, whose proofs are given in Section 3.7.

Proposition 3.4.12. *Assume that the pair* $(\overline{h}_1, \overline{l}_1)$ *is as in* (3.7.10). *Then we have*

$$\begin{aligned} V_1(T) &= V_0'(T), && (3.4.70)\\ V_3(T) &= -3\hat{A}'(T)V_0'(T), && (3.4.71)\\ V_5(T) &= 15\hat{A}'(T)^2 V_0'(T), && (3.4.72)\\ V_7(T) &= -105\hat{A}'(T)^3 V_0'(T). && (3.4.73) \end{aligned}$$

Proposition 3.4.13. *Assume that the pair* $(\overline{h}_1, \overline{l}_1)$ *is as in* (3.7.10). *Then we have*

$$\begin{aligned} U_2(T) &= -D'(T)U_0(T) - 2V_0'(T), && (3.4.74)\\ U_4(T) &= \big(D'(T) + 2\hat{A}'(T)\big)V_0'(T), && (3.4.75)\\ U_6(T) &= -3(8D'(T) + 11\hat{A}'(T))\hat{A}'(T)V_0'(T), && (3.4.76)\\ U_8(T) &= 6\big(17D'(T) + 64\hat{A}'(T)\big)\hat{A}'(T)^2 V_0'(T). && (3.4.77) \end{aligned}$$

□

3.5 Examples

This section is devoted to examples of vehicles with "quite simple" shapes, for which the coefficients in the matrices in (3.3.23)-(3.3.30) can be be computed explicitly. We begin with the example of a vehicle with one axis of revolution, for which the controllability fails for any choice of the flow controls.

3.5.1 Solid of revolution

Let $f \in \mathcal{C}^1([a,b];\mathbb{R})$ be a nonnegative function such that $f(a) = f(b) = 0$, and let

$$S = \Big\{\Big(y_1, \delta f(y_1)\cos(\beta), \delta f(y_1)\sin(\beta)\Big);\ y_1 \in [a,b],\ \delta \in [0,1],\ \beta \in [0,2\pi)\Big\}$$

Assume that the density ρ depends on y_1 only, i.e. $\rho = \rho(y_1)$. Clearly $J_0 = \mathrm{diag}(J_1, J_2, J_2)$. On the other hand,

$$\partial\Omega = \Big\{\Big(y_1, f(y_1)\cos(\beta), f(y_1)\sin(\beta)\Big)\ :\ y_1 \in [a,b],\ \beta \in [0,2\pi)\Big\},$$

and the normal vector $\hat{n}$ to $\partial\Omega$ is given by

$$\hat{n}(y_1,\beta) = \frac{1}{\sqrt{1 + (f'(y_1))^2}}\Big(-f'(y_1), \cos(\beta), \sin(\beta)\Big)^*,$$

so that

$$(y \times \hat{n})(y_1,\beta) = \frac{(y_1 + f(y_1)f'(y_1))}{\sqrt{1 + (f'(y_1))^2}}\Big(0, -\sin(\beta), \cos(\beta)\Big)^*.$$

It follows that $(y \times \hat{n})e_1 = 0$. Replacing in (3.2.23), we obtain

$$J_1\dot{r}_1 = (J_0\dot{r})e_1 = -(r \times J_0 r)e_1 = J_2 r_2 r_3 - J_2 r_2 r_3 = 0,$$

which indicates that the angular velocity r_1 is not controllable.

3.5.2 Ellipsoidal vehicle.

We assume here that the vehicle fills the ellipsoid

$$S = \Big\{ y \in \mathbb{R}^3; \quad (y_1/c_1)^2 + (y_2/c_2)^2 + (y_3/c_3)^2 \leq 1 \}$$

where $c_1 > c_2 > c_3 > 0$ denote some numbers. Our first aim is to compute explicitly the functions ϕ_i and φ_i for $i = 1, 2, 3$ which solve (3.3.20)-(3.3.22) for

$$\Omega = \Big\{ y \in \mathbb{R}^3; \quad (y_1/c_1)^2 + (y_2/c_2)^2 + (y_3/c_3)^2 > 1 \}.$$

Computations of the functions ϕ_i and φ_i.

We follow closely [60, pp.148-155]. We introduce a special system of orthogonal curvilinear coordinates, denoted by (λ, μ, ν), which are defined as the roots of the equation

$$\frac{y_1^2}{c_1^2 + \theta} + \frac{y_2^2}{c_2^2 + \theta} + \frac{y_3^2}{c_3^2 + \theta} - 1 = 0, \tag{3.5.1}$$

viewed as a cubic in θ. It is clear that (3.5.1) has three real roots: $\lambda \in (-c_3^2, +\infty)$, $\mu \in (-c_2^2, -c_3^2)$, and $\nu \in (-c_1^2, -c_2^2)$.

It follows immediately from the above definition of λ, μ, ν, that

$$\frac{y_1^2}{c_1^2 + \theta} + \frac{y_2^2}{c_2^2 + \theta} + \frac{y_3^2}{c_3^2 + \theta} - 1 = \frac{(\lambda - \theta)(\mu - \theta)(\nu - \theta)}{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}.$$

This yields

$$\begin{aligned}
y_1^2 &= \frac{(c_1^2 + \lambda)(c_1^2 + \mu)(c_1^2 + \nu)}{(c_2^2 - c_1^2)(c_3^2 - c_1^2)}, & \partial_\lambda y_1 &= \frac{1}{2}\frac{y_1}{(c_1^2 + \lambda)}, \\
y_2^2 &= \frac{(c_2^2 + \lambda)(c_2^2 + \mu)(c_2^2 + \nu)}{(c_1^2 - c_2^2)(c_3^2 - c_2^2)}, & \partial_\lambda y_2 &= \frac{1}{2}\frac{y_2}{(c_2^2 + \lambda)}, \\
y_3^2 &= \frac{(c_3^2 + \lambda)(c_3^2 + \mu)(c_3^2 + \nu)}{(c_1^2 - c_3^2)(c_2^2 - c_3^2)}, & \partial_\lambda y_3 &= \frac{1}{2}\frac{y_3}{(c_3^2 + \lambda)}.
\end{aligned} \tag{3.5.2}$$

We introduce the scale factors

$$\begin{aligned}
h_\lambda &= \frac{1}{2}\sqrt{\frac{(\lambda - \mu)(\lambda - \nu)}{(\lambda + c_1^2)(\lambda + c_2^2)(\lambda + c_3^2)}}, \\
h_\mu &= \frac{1}{2}\sqrt{\frac{(\mu - \nu)(\mu - \lambda)}{(\mu + c_1^2)(\mu + c_2^2)(\mu + c_3^2)}}, \\
h_\nu &= \frac{1}{2}\sqrt{\frac{(\nu - \lambda)(\nu - \mu)}{(\nu + c_1^2)(\nu + c_2^2)(\nu + c_3^2)}},
\end{aligned} \tag{3.5.3}$$

and the function

$$f(x) = \sqrt{(y_1 + c_1^2)(y_2 + c_2^2)(y_3 + c_3^2)}.$$

If ξ is any smooth function of λ, then its Laplacian is given by

$$\Delta\xi = \frac{4}{(\lambda-\mu)(\lambda-\nu)} f(\lambda)\partial_\lambda(f(\lambda)\partial_\lambda\xi). \tag{3.5.4}$$

according to [60, (7) p. 150]. We search ϕ_i in the form $\phi_i(y_1, y_2, y_3) = y_i\xi_i(y_1, y_2, y_3)$. Then

$$0 = \Delta\phi_i = y_i\Delta\xi_i + 2\partial_i\xi_i, \tag{3.5.5}$$

Assuming in addition that ξ_i depends only on λ, then we obtain

$$\frac{2\partial_i\xi_i}{y_i} = \frac{2\partial_\lambda y_i}{y_i}\frac{\partial_\lambda\xi_i}{h_\lambda^2} = \frac{1}{c_i^2+\lambda}\frac{\partial_\lambda\xi_i}{h_\lambda^2} = \frac{4f^2(\lambda)}{(c_i^2+\lambda)}\frac{\partial_\lambda\xi_i}{(\lambda-\mu)(\lambda-\nu)}. \tag{3.5.6}$$

Combining (3.5.5) to (3.5.4) and (3.5.6), we arrive to

$$0 = \partial_\lambda(f(\lambda)\partial_\lambda\xi_i) + \frac{1}{c_i^2+\lambda} f(\lambda)\partial_\lambda\xi_i,$$

which is readily integrated in

$$\xi_i = -C_i^{te}\int_\lambda^{+\infty}\frac{dx}{(c_i^2+x)f(x)} + C^{te}.$$

We choose the constant $C^{te} = 0$ for (3.3.22) to be fulfilled. As $\partial\Omega$ is represented by the equation $\lambda = 0$, then (3.3.21) reads

$$\partial_{\hat{n}}\phi_i = \hat{n}_i \Leftrightarrow \xi_i\frac{\partial_\lambda y_i}{y_i} + \partial_\lambda\xi_i = \frac{\partial_\lambda y_i}{y_i}.$$

We infer that $C_i^{te} = c_1c_2c_3/(2-\alpha_i)$, where

$$\alpha_i = c_1c_2c_3\int_0^{+\infty}\frac{dx}{(c_i^2+x)f(x)}.$$

It is easy seen that

$$\frac{2c_2c_3}{3c_1^2} \le \alpha_i \le \frac{2c_1c_2}{3c_3^2}.$$

It follows that if c_1, c_2, c_3 are sufficiently close, then α_i is different from 2, so that C_i^{te} is well defined. We conclude that

$$\phi_i(y) = -\frac{\alpha_i}{2-\alpha_i}y_i, \quad \forall y \in \partial\Omega. \tag{3.5.7}$$

Let us now proceed to the computation of φ_i. We search φ_i in the form $\varphi_i(y) = \frac{y_1y_2y_3}{y_i}\xi_i(y)$, where ξ_i depends only on λ. We obtain

$$\Delta\xi_i + 2\sum_{j=1, j\neq i}^{3}\frac{\partial_{y_j}\xi_i}{y_j} = 0 \Leftrightarrow \partial_\lambda(f(\lambda)\partial_\lambda\xi_i) + \Big(\sum_{j=1,j\neq i}^{3}\frac{1}{(c_j^2+\lambda)}\Big)f(\lambda)\partial_\lambda\xi_i = 0,$$

and hence

$$\xi_i = -C_i^{te} \int_{\lambda}^{+\infty} \frac{(c_i^2 + x)}{f^3(x)} dx.$$

From (3.3.21)-(3.3.22), we infer that

$$C_1^{te} = \frac{c_1 c_2 c_3 (c_2^2 - c_3^2)}{2 - \beta_1}, \qquad \beta_1 = c_1 c_2 c_3 (c_2^2 + c_3^2) \int_0^{+\infty} \frac{dx}{(c_2^2 + x)(c_3^2 + x) f(x)},$$

$$C_2^{te} = \frac{c_1 c_2 c_3 (c_3^2 - c_1^2)}{2 - \beta_2}, \qquad \beta_2 = c_1 c_2 c_3 (c_3^2 + c_1^2) \int_0^{+\infty} \frac{dx}{(c_3^2 + x)(c_1^2 + x) f(x)},$$

$$C_3^{te} = \frac{c_1 c_2 c_3 (c_1^2 - c_2^2)}{2 - \beta_3}, \qquad \beta_3 = c_1 c_2 c_3 (c_1^2 + c_2^2) \int_0^{+\infty} \frac{dx}{(c_1^2 + x)(c_2^2 + x) f(x)}.$$

Note that at the limit case $c_1 = c_2 = c_3$, we obtain $\beta_1 = \beta_2 = \beta_3 = 4/5$. Therefore, if c_1, c_2 and c_3 are near but different, then β_i is different from 2, and therefore C_i^{te} is well defined. We conclude that

$$\varphi_i = -\left(C_i^{te} \int_0^{+\infty} \frac{(c_i^2 + x)}{f^3(x)} dx \right) \frac{y_1 y_2 y_3}{y_i}, \ \forall y \in \partial\Omega. \tag{3.5.8}$$

Controllability of the ellipsoid with six controls

Assume still that S is given by (3.5.2). Note that S is symmetric with respect to the plane $\{y_p = 0\}$ for $p = 1, 2, 3$. Assume given six functions χ_j, $j = 1, .., 6$, each being symmetric with respect to the planes $\{y_p = 0\}$ for $p = 1, 2, 3$, with

$$\begin{gathered}
\varepsilon_{\chi_1}^p = \begin{cases} -1 & p = 1 \\ 1 & p = 2 \\ 1 & p = 3 \end{cases}, \quad \varepsilon_{\chi_2}^p = \begin{cases} 1 & p = 1 \\ -1 & p = 2 \\ 1 & p = 3 \end{cases}, \quad \varepsilon_{\chi_3}^p = \begin{cases} 1 & p = 1 \\ 1 & p = 2 \\ -1 & p = 3 \end{cases}, \\
\varepsilon_{\chi_4}^p = \begin{cases} 1 & p = 1 \\ -1 & p = 2 \\ -1 & p = 3 \end{cases}, \quad \varepsilon_{\chi_5}^p = \begin{cases} -1 & p = 1 \\ 1 & p = 2 \\ -1 & p = 3 \end{cases}, \quad \varepsilon_{\chi_6}^p = \begin{cases} -1 & p = 1 \\ -1 & p = 2 \\ 1 & p = 3 \end{cases}.
\end{gathered} \tag{3.5.9}$$

We notice that C is a diagonal matrix $C = -\text{diag}(C_1, C_2, C_3, C_4, C_5, C_6)$, with

$$C_i = \int_{\partial\Omega} \phi_i \chi_i, \ i = 1, 2, 3, \quad \text{and} \quad C_{i+3} = \int_{\partial\Omega} \varphi_i \chi_{i+3}, \ i = 1, 2, 3.$$

From (3.5.7)-(3.5.8), there are some constants $C_i^{te} \neq 0$, $i = 1, ..., 6$, which depend only on c_1, c_2, and c_3, and such that

$$C_i = C_i^{te} \int_{\{y_i > 0\} \cap \partial\Omega} y_i \chi_i(y), \quad C_{i+3} = C_{i+3}^{te} \int_{\{\frac{y_1 y_2 y_3}{y_i} > 0\} \cap \partial\Omega} \left(\frac{y_1 y_2 y_3}{y_i} \right) \chi_{i+3}(y), \quad i = 1, 2, 3. \tag{3.5.10}$$

By (3.5.10), we have that $C_i \neq 0$ for $i = 1, ..., 6$, and hence $\text{rank}(C) = 6$ if, in addition to (3.5.9), it holds

$$\chi_i \not\equiv 0, \ i = 1, ..., 6 \tag{3.5.11}$$
$$\chi_i \geq 0 \text{ on } \{y_i > 0\}, \ i = 1, 2, 3, \quad \chi_{i+3} \geq 0 \text{ on } \{\frac{y_1 y_2 y_3}{y_i} > 0\}, \ i = 1, 2, 3. \tag{3.5.12}$$

By Proposition 3.4.1 and Theorem 3.4.9, it follows that both the linearized system (3.4.1) and the nonlinear system (3.3.67) are controllable.

Controllability of the ellipsoid with four controls

We consider the same controllers χ_1, χ_4, χ_5 and χ_6 as above (satisfying (3.5.9), (3.5.11), (3.5.12)). If we let $m \to +\infty$ and $J_0 \to +\infty$, then the matrices $B_\infty = \lim_{m, J_0 \to +\infty} B$ and C are given by

$$C = -\begin{pmatrix} C_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & C_4 & 0 & 0 \\ 0 & 0 & C_5 & 0 \\ 0 & 0 & 0 & C_6 \end{pmatrix}, \quad B_\infty = -\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & B_6 \\ 0 & 0 & B_5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

with

$$B_5 = \int_{\partial\Omega} (\nabla\psi_1 \cdot \nabla\psi_5)\, \hat{n}_3, \quad B_6 = \int_{\partial\Omega} (\nabla\psi_1 \cdot \nabla\psi_6)\, \hat{n}_2.$$

Thus, if $B_5 \neq 0$ and $B_6 \neq 0$, we see that (3.4.66) and (3.4.67) are fulfilled, so that the local controllability of (3.3.67) is ensured by Corollary 3.4.10 for m and J_0 large enough.

Controllability of the ellipsoid with three controls

Assuming that χ_1, χ_4, χ_5 and χ_6 are as above (satisfying (3.5.9), (3.5.11), (3.5.12)), we consider the controls supported by χ_1, χ_4 and $\tilde{\chi}_5 = \chi_5 + \chi_6$.

If we let $\alpha \to 0$ (without assuming that $m \to +\infty$ and $J_0 \to +\infty$), then the matrices $B_\infty = \lim_{m, J_0 \to +\infty} B$ and C take the form

$$C = -\begin{pmatrix} C_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & C_4 & 0 \\ 0 & 0 & C_5 \\ 0 & 0 & C_6 \end{pmatrix}, \quad B_\infty = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & B_6 \\ 0 & 0 & B_5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

Let K denote the 4×4 matrix obtained from A by removing the first and fourth lines (resp. columns), and let b the vector in $\mathbb{R}^4$ obtained from E by removing the first and fourth coordinates. Then (3.4.68) and (3.4.69) are satisfied for α small enough whenever

$$\text{rank } (b, Kb, K^2b, K^3b) = 4. \tag{3.5.13}$$

Note that (3.5.13) is nothing else than the classical Kalman rank test for the system $\dot{x} = Kx + bu$, and that this rank condition is satisfied if e.g. the eigenvalues of K are pairwise distinct and b has a nonzero component along each eigenspace.

3.6 Quaternions and rotations

Quaternions are a convenient tool for representing rotations of objects in three dimensions. For that reason, they are widely used in robotic, navigation, flight dynamics, etc. (See e.g. [2, 88]). We limit ourselves to introducing the few definitions and properties needed to deal with the dynamics of h and Q. (We refer the reader to [2] for more details.)

The set of quaternions, denoted by $\mathbb{H}$, is a noncommutative field containing $\mathbb{C}$ and which is a $\mathbb{R}$-algebra of dimension 4. Any quaternion $q \in \mathbb{H}$ may be written as

$$q = q_0 + q_1 i + q_2 j + q_3 k,$$

where $(q_0, q_1, q_2, q_3) \in \mathbb{R}^4$ and $i, j, k \in \mathbb{H}$ are some quaternions whose products will be given later. We say that q_0 (resp. $q_1 i + q_2 j + q_3 k$) is the *real part* (resp. the *imaginary part*) of q. Identifying the imaginary part $q_1 i + q_2 j + q_3 k$ with the vector $\vec{q} = (q_1, q_2, q_3) \in \mathbb{R}^3$, we can represent the quaternion q as $q = [q_0, \vec{q}]$, where $q_0 \in \mathbb{R}$ (resp. $\vec{q} \in \mathbb{R}^3$) is the *scalar part* (resp. the *vector part*) of q. The addition, scalar multiplication and quaternion multiplication are defined respectively by

$$\begin{aligned}
&[p_0, \vec{p}] + [q, \vec{q}] = [p_0 + q_0, \vec{p} + \vec{q}],\\
&t[q_0, \vec{q}] = [tq_0, t\vec{q}],\\
&[p_0, \vec{p}] * [q_0, \vec{q}] = [p_0 q_0 - \vec{p} \cdot \vec{q}, p_0 \vec{q} + q_0 \vec{p} + \vec{p} \times \vec{q}],
\end{aligned}$$

where "$\cdot$" is the dot product and "$\times$" is the cross product. We stress that the quaternion multiplication $*$ is not commutative. Actually, we have that

$$\begin{aligned}
&i * j = k, \quad j * i = -k,\\
&j * k = i, \quad k * j = -i,\\
&k * i = j, \quad i * k = -j,\\
&i^2 = j^2 = k^2 = -1.
\end{aligned}$$

Any pure scalar q_0 and any pure vector $\vec{q}$ may be viewed as quaternions

$$q_0 = [q_0, \vec{0}\,], \qquad \vec{q} = [0, \vec{q}],$$

and hence any quaternion $q = [q_0, \vec{q}]$ can be written as the sum of a scalar and a vector, namely

$$q = q_0 + \vec{q}.$$

The cross product of vectors extends to quaternions by setting

$$p \times q = \frac{1}{2}(p * q - q * p) = [0, \vec{p} \times \vec{q}].$$

The *conjugate* of a quaternion $q = [q_0, \vec{q}]$ is $q^* = [q_0, -\vec{q}]$. The *norm* of q is

$$||q|| = (|q_0|^2 + ||\vec{q}||^2)^{\frac{1}{2}}.$$

From

$$q * q^* = q^* * q = ||q||^2,$$

we infer that

$$q^{-1} = \frac{q^*}{||q||^2}.$$

A *unit quaternion* is a quaternion of norm 1. The set of unit quaternions may be identified with S^3. It is a group for $*$.

Any unit quaternion $q = [q_0, \vec{q}]$can be written in the form

$$q = \cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\vec{u}, \tag{3.6.1}$$

where $\alpha \in \mathbb{R}$, and $\vec{u} \in \mathbb{R}^3$ with $||\vec{u}|| = 1$. Note that the writing is not unique: if the pair $(\alpha, \vec{u}\,)$ is convenient, the same is true for the pairs $(-\alpha, -\vec{u}\,)$ and $(\alpha + 4k\pi, \vec{u}\,)$ $(k \in \mathbb{Z})$, as well. However, if we impose that $\alpha \in [0, 2\pi]$, then α is unique, and $\vec{u}$ is unique for $|q_0| < 1$. (However, any $\vec{u} \in S^3$ is convenient for $|q_0 = 1$.)

For any unit quaternion q, let the matrix $R(q) \in \mathbb{R}^{3\times 3}$ be defined by

$$R(q)v = q * v * q^* \quad \forall v \in \mathbb{R}^3. \tag{3.6.2}$$

Then $R(q)$ is found to be

$$R(q) = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_2q_1 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_3q_1 - q_0q_2) & 2(q_3q_2 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}$$

For q given by (3.6.1), then $R(q)$ is the *rotation around the axis* $\mathbb{R}\vec{u}$ *of angle* α.

Note that $R(q_1 * q_2) = R(q_1)R(q_2)$ (i.e. R is a morphism of groups), hence

$$R(1) = Id, \quad R(q)^{-1} = R(q^*).$$

We notice that the map $q \to R(q)$ from the unit quaternions set S^3 to $SO(3)$ is onto, but not one-to-one, for $R(-q) = R(q)$. It becomes one-to-one when restricted to the open set

$$S^3_+ := \{q = [q_0, \vec{q}] \in \mathbb{H};\ ||q|| = 1 \text{ and } q_0 > 0\}.$$

Furthermore, the map R is a smooth invertible map from S^3_+ onto an open neighbourhood $\mathcal{O}$ of Id in $SO(3)$. On the other hand, the map

$$\vec{q} \to q = [q_0, \vec{q}] = [\sqrt{1 - ||\vec{q}||^2}, \vec{q}]$$

is a smooth invertible map from the unit ball $B_1(0) = \{\vec{q} \in \mathbb{R}^3;\ ||\vec{q}|| < 1\}$ onto S^3_+. Thus the rotations in $\mathcal{O}$ can be parameterized by $\vec{q} \in B_1(0)$.

3.7 Proof of some propositions

In this section we devote in the proofs of the Propositions 3.4.8 and 3.4.12.

3.7.1 Proof of Proposition 3.4.8.

Let us prove by induction on $k \in \mathbb{N}$ that

$$V_{2k}^{(2l)}(T) = 0 \qquad \forall l \in \mathbb{N}. \tag{3.7.1}$$

The property is clearly true for $k = 0$, since

$$V_0^{(2l)}(T) = \hat{B}^{(2l)}(T) + \hat{A}^{(2l)}(T)\hat{C} = 0,$$

by (3.4.54). Assume that (3.7.1) is established for some $k \in \mathbb{N}$. Then by (3.4.53) applied twice, we have

$$V_{2k+2} = V_{2k}'' - 2\hat{A}V_{2k}' - A'V_{2k} + A^2V_{2k},$$

hence

$$V_{2k+2}^{(2l)}(T) = V_{2k}^{(2l+2)} - 2(AV_{2k}')^{(2l)}(T) - (A'V_{2k})^{(2l)}(T) + (A^2V_{2k})^{(2l)}(T). \tag{3.7.2}$$

The first term in the r.h.s. of (3.7.2) is null by (3.7.1). The second one is also null, for by Leibniz' rule

$$(AV_{2k}')^{(2l)}(T) = \sum_{p=0}^{2l} C_{2l}^p A^{(p)}(T) V_{2k}^{(2l-p+1)}(T)$$

and $A^{(p)}(T) = 0$ if p is even, while $V_{2k}^{(2l-p+1)}(T) = 0$ if p is odd. One proves in a similar way that the third and fourth terms in the r.h.s. of (3.7.2) are null, noticing that for p odd we have

$$(A^2)^{(p)}(T) = 2(AA')^{(p-1)}(T) = 0. \tag{3.7.3}$$

From (3.7.1), we infer that

$$V_{2k+1}^{(2l+1)}(T) = V_{2k}^{(2l+2)}(T) - (AV_{2k})^{(2l+1)}(T) = 0.$$

Let us proceed to the proof of (3.4.56). Again, we first prove by induction on $k \in \mathbb{N}$ that

$$U_{2k+1}^{(2l)}(T) = 0 \qquad \forall l \in \mathbb{N}. \tag{3.7.4}$$

It follows from (3.4.52)-(3.4.53) and (3.4.55) that

$$U_1^{(2l)}(T) = U_0^{(2l+1)}(T) - (DU_0)^{(2l)}(T) - V_0^{(2l)}(T) = 0 \qquad k \in \mathbb{N}.$$

Assume that (3.7.4) is true for some $k \in \mathbb{N}$. Then, by (3.4.53) applied twice,

$$U_{2k+3}^{(2l)}(T) = U_{2k+1}^{(2l+2)} - (Du_{2k+1})^{(2l+1)}(T) - V_{2k+1}^{(2l+1)} - (DU_{2k+2})^{(2l)}(T) - V_{2k+2}^{(2l)}(T). \tag{3.7.5}$$

Using (3.4.54), (3.4.55) and (3.7.4), we see that all the terms in the r.h.s. of (3.7.5), except possibly $(DU_{2k+2})^{(2l)}(T)$, are null. Finally,

$$(DU_{2k+2})^{(2l)}(T) = (DU'_{2k+1})^{(2l)}(T) - (D^2U_{2k+1})^{(2l)}(T) - (DV_{2k+1})^{(2l)}(T).$$

Using Leibniz' rule for each term, noticing that in each pair (p,q) with $p+q=2l$, p and q are simultaneously even or odd, and using (3.4.54)-(3.4.55), (3.7.3) (with $\hat{A}$ replaced by D), and (3.7.4), we conclude that $(DU_{2k+2})^{(2l)}(T)=0$, so that $U^{(2l)}_{2k+3}(T)=0$.

Finally, $U^{(2l+1)}_{2k}(T)=0$ is obvious for $k=0$, while for $k\geq 1$

$$U^{(2l+1)}_{2k}(T) = U^{(2l+2)}_{2k-1}(T) - (DU_{2k-1})^{(2l+1)}(T) - V^{(2l+1)}_{2k-1}(T) = 0$$

by (3.4.54), (3.4.55) and (3.7.4) (with $2k+1$ replaced by $2k-1$). The proof of Proposition 3.4.8 is complete.

3.7.2 Proof of Proposition 3.4.12

From (3.4.51),(3.4.52) and (3.4.37), we obtain successively

$$\begin{aligned}
V_1(T) &= V_0'(T) = \hat{B}'(T) + \hat{A}'(T)\hat{C} = \overline{w_1}'(T)\big(\mathcal{I}^{-1}B + \mathcal{I}^{-1}A\mathcal{I}^{-1}C\big)\\
V_3(T) &= V_2'(T)\\
&= (V_1' - \hat{V}_1)'(T)\\
&= (V_0' - \hat{A}V_0)''(T) - (\hat{A}V_1)'(T)\\
&= V_0'''(T) - 2\hat{A}'(T)V_0'(T) - \hat{A}'(T)V_1(T)\\
&= -3\hat{A}'(T)V_0'(T).
\end{aligned}$$

Successive applications of (3.4.52) yield

$$V_5(T) = V_0^{(5)}(T) - \sum_{i=0}^{3}(\hat{A}V_i)^{(4-i)}(T), \tag{3.7.6}$$

$$V_7(T) = V_0^{(7)}(T) - \sum_{i=0}^{5}(\hat{A}V_i)^{(6-i)}(T). \tag{3.7.7}$$

Since $V_0^{(k)}(T)=0$ for $k\geq 2$, it remains to estimate the terms $(\hat{A}V_i)^{(4-i)}(T)$ and $(\hat{A}V_i)^{(6-i)}(T)$. Notice first that by (3.4.37) and Leibniz' rule

$$(\hat{A}V_i)^{(k)}(T) = k\hat{A}'(T)V_i^{(k-1)}(T).$$

Thus, from (3.4.37) and (3.4.71), we have that

$$(\hat{A}V_0)^{(4)}(T) = 0, \tag{3.7.8}$$

$$(\hat{A}V_1)^{(3)}(T) = 3\hat{A}'(T)V_1''(T) = 3\hat{A}'(T)\big(V_0^{(3)}(T) - (\hat{A}V_0)''(T)\big) = -6\hat{A}'(T)^2V_0'(T), \tag{3.7.9}$$

$$(\hat{A}V_2)''(T) = 2\hat{A}'(T)V_2'(T) = 2\hat{A}'(T)V_3(T) = -6\hat{A}'(T)^2V_0'(T), \tag{3.7.10}$$

$$(\hat{A}V_3)'(T) = \hat{A}'(T)V_3(T) = -3\hat{A}'(T)^2V_0'(T). \tag{3.7.11}$$

This yields (3.4.72). On the other hand

$$
\begin{aligned}
(\hat{A}V_0)^{(6)}(T) &= 0, && (3.7.12)\\
(\hat{A}V_1)^{(5)}(T) &= 5\hat{A}'(T)V_1^{(4)}(T) = 5\hat{A}'(T)\big(V_0^{(5)} - (\hat{A}V_0)^{(4)}\big)(T) = 0, && (3.7.13)\\
(\hat{A}V_2)^{(4)}(T) &= 4\hat{A}'(T)V_2^{(3)}(T). && (3.7.14)
\end{aligned}
$$

Since

$$V_2 = V_1' - \hat{A}V_1 = V_0'' - (\hat{A}V_0)' - \hat{A}V_1,$$

we obtain with (3.4.37) and (3.7.9) that

$$V_2^{(3)}(T) = V_0^{(5)}(T) - (\hat{A}V_0)^{(4)}(T) - (\hat{A}V_1)^{(3)}(T) = 6\hat{A}'(T)^2 V_0'(T),$$

hence

$$(\hat{A}V_2)^{(4)}(T) = 24\hat{A}'(T)^3 V_0'(T). \tag{3.7.15}$$

On the other hand,

$$
\begin{aligned}
(\hat{A}V_3)^{(3)}(T) &= 3\hat{A}'(T)V_3''(T)\\
&= 3\hat{A}'(T)\big(V_4'(T) + (\hat{A}V_3)'(T)\big)\\
&= 3\hat{A}'(T)\big(V_5(T) + \hat{A}'(T)V_3(T)\big)\\
&= 36\hat{A}'(T)^3 V_0'(T) && (3.7.16)
\end{aligned}
$$

where we used (3.4.52) and (3.4.71)-(3.4.72). Finally,

$$(\hat{A}V_4)''(T) = 2\hat{A}'(T)V_4'(T) = 2\hat{A}'(T)V_5(T) = 30\hat{A}'(T)^3 V_0'(T) \tag{3.7.17}$$

and

$$(\hat{A}V_5)'(T) = \hat{A}'(T)V_5(T) = 15\hat{A}'(T)V_0'(T). \tag{3.7.18}$$

Gathering together (3.7.7) and (3.7.12)-(3.7.18), we obtain (3.4.73). The proof of Proposition 3.4.12 is complete.

3.7.3 Proof of Proposition 3.4.13

From (3.4.52)-(3.4.53), we have that

$$U_0 \equiv \hat{C}, \qquad U_i = U_{i-1}' - DU_{i-1} - V_{i-1}, \quad \forall i \geq 1. \tag{3.7.19}$$

Thus

$$
\begin{aligned}
U_2(T) &= (U_1' - DU_1 - V_1)(T)\\
&= (0 - (DU_0)' - V_0')(T) - V_1(T)\\
&= -D'(T)U_0 - 2V_0'(T)
\end{aligned}
$$

where we used successively (3.7.19), (3.4.56) and (3.4.70).

Successive applications of (3.7.19) yield

$$U_4(T) = -\sum_{i=0}^{3}[(DU_i)^{(3-i)} + V_i^{(3-i)}](T). \tag{3.7.20}$$

Using (3.4.37), we obtain that

$$\begin{aligned}\sum_{i=0}^{3}(DU_i)^{(3-i)}(T) &= \sum_{i=0}^{2}(3-i)D'(T)U_i^{(2-i)}(T) \\ &= 2D'(T)\big(U_2(T)+V_1(T)\big)+2D'(T)V_0'(T) \\ &= -3D'(T)\big(D'(T)U_0+2V_0'(T)\big)+2D'(T)V_0'(T) \\ &= -4D'(T)V_0'(T)\end{aligned} \tag{3.7.21}$$

where we used (3.4.70), (3.4.74) and the fact that $D'(T)^2 = 0$.

On the other hand,

$$\begin{aligned}\sum_{i=0}^{3}V_i^{(3-i)}(T) &= (V_0' - \hat{A}V_0)''(T) + V_2'(T) + V_3(T) \\ &= -2\hat{A}'(T)V_0'(T) + 2V_3(T) \\ &- -8\hat{A}'(T)V_0'(T)\end{aligned} \tag{3.7.22}$$

by (3.4.71). Combining (3.7.20)-(3.7.22), we obtain (3.4.75).

Let us now compute $U_6(T)$. Successive applications of (3.7.19) yield

$$U_6(T) = -\sum_{i=0}^{5}[(DU_i)^{(5-i)} + V_i^{(5-i)}](T). \tag{3.7.23}$$

We have that

$$\sum_{i=0}^{5}(DU_i)^{(5-i)}(T) = \sum_{i=0}^{4}(5-i)D'(T)U_i^{(4-i)}(T).$$

Let us estimate the terms $U_i^{(4-i)}(T)$ for $i = 0, ..., 4$ Obviously, $U_0^{(4)}(T) = 0$ by (3.7.19), while by (3.4.37)

$$U_1^{(3)}(T) = -(DU_0)^{(3)}(T) - V_0^{(3)}(T) = 0. \tag{3.7.24}$$

Next we use (3.7.19) to obtain successively

$$\begin{aligned}U_3'(T) &= U_4(T) + V_3(T), \qquad (3.7.25) \\ U_2''(T) &= U_3'(T) + (DU_2)'(T) + V_2'(T), \\ &= U_4(T) + V_3(T) + D'(T)U_2(T) + V_3(T).\end{aligned}$$

It follows that

$$\begin{aligned}&\sum_{i=0}^{4}(DU_i)^{(5-i)}(T) \\ &= 3D'(T)\big(U_4(T)+2V_3(T)+D'(T)U_2(T)\big)+2D'(T)\big(U_4(T)+V_3(T)\big)+D'(T)U_4(T), \\ &= D'(T)\big(6U_4(T)+8V_3(T)\big), \\ &= 24D'(T)(D'(T)+2\hat{A}'(T))V_0'(T) - 24D'(T)\hat{A}'(T)V_0'(T), \\ &= 24D'(T)\hat{A}'(T)V_0'(T).\end{aligned} \tag{3.7.26}$$

On the other hand, using (3.7.13)-(3.7.16) and (3.4.71)-(3.4.72), we have

$$\begin{aligned}
\sum_{i=0}^{4} V_i^{(5-i)}(T) &= V_1^{(4)}(T + V_2^{(3)}(T) + V_3^{(2)}(T) + V_4'(T) && (3.7.27)\\
&= 6\hat{A}'(T)^2 V_0'(T) + 2V_5(T) + \hat{A}'(T)V)3(T) && (3.7.28)\\
&= 33\hat{A}'(T)^2 V_0'(T) && (3.7.29)
\end{aligned}$$

(3.4.76) follows from (3.7.23)-(3.7.29).

Finally, we compute $U_8(T)$. We see that

$$U_8(T) = -\sum_{i=0}^{7}[(DU_i)^{(7-i)} + V_i^{(7-i)}](T). \tag{3.7.30}$$

Then

$$\begin{aligned}
\sum_{i=0}^{7}(DU_i)^{(7-i)}(T) &= \sum_{i=0}^{6}(7-i)D'(T)U_i^{(6-i)}(T)\\
&= 6D'(T)U_1^{(5)}(T) + 5D'(T)U_2^{(4)}(T) + 4D'(T)U_3^{(3)}(T)\\
&\quad +3D'(T)U_4''(T) + 2D'(T)U_5'(T) + D'(T)U_6(T).
\end{aligned}$$

Using (3.4.37) and (3.7.19), we readily see that

$$U_1^{(5)}(T) = U_2^{(4)}(T) = 0.$$

Next, successive applications of (3.7.19) give

$$\begin{aligned}
U_5'(T) &= U_6(T) + V_5(T),\\
U_4''(T) &= U_5'(T) + (DU_4)'(T) + V_4'(T),\\
&= U_6(T) + D'(T)U_4(T) + 2V_5(T).\\
U_3^{(3)}(T) &= U_4''(T) + (DU_3)''(T) + V_3''(T),\\
&= \big(U_6(T) + D'(T)U_4(T) + 2V_5(T)\big) + 2D'(T)\big(U_4(T) + V_3(T)\big),\\
&\quad +V_5(T) + \hat{A}'(T)V_3(T).
\end{aligned}$$

Thus

$$\begin{aligned}
\sum_{i=0}^{7}(DU_i)^{(7-i)}(T) &= 4D'(T)\big(U_6(T) + 3D'(T)U_4(T) + 2V_5(T) + 2D'(T)V_3(T) + \hat{A}'(T)V_3(T)\big)\\
&\quad +3D'(T)\big(U_6(T) + D'(T)U_4(T) + 2V_5(T)\big)\\
&\quad +2D'(T)\big(U_6(T) + V_5(T)\big) + D'(T)U_6(T),\\
&= D'(T)[10\,U_6(T) + 16\,V_5(T) + 4\hat{A}'(T)V_3(T)],\\
&= D'(T)[-240\,D'(T)\hat{A}'(T)V_0'(T) - 330\,\hat{A}'(T)V_0'(T)\\
&\quad +240\,\hat{A}'(T)^2V_0'(T) - 12\hat{A}'(T)^2V_0'(T)]\\
&= -102\,D'(T)\hat{A}'(T)^2V_0'(T). && (3.7.31)
\end{aligned}$$

It remains to compute $\sum_{i=0}^{7} V_i^{(7-i)}(T)$. It is easy to see that

$$V_0^{(7)}(T) = V_1^{(6)}(T) = V_2^{(5)}(T) = 0.$$

Successive applications of (3.4.52) give

$$\begin{aligned}
V_6'(T) &= V_7(T) \\
V_5''(T) &= V_6'(T) + (\hat{A}V_5)'(T) = V_7(T) + \hat{A}'(T)V_5(T) \\
V_4^{(3)} &= V_5''(T) + (\hat{A}V_4)''(T) \\
&= V_7(T) + 3\hat{A}'(T)V_5(T) \\
V_3^{(4)}(T) &= V_4^{(3)}(T) + (\hat{A}V_3)^{(3)}(T) \\
&= V_7(T) + 3\hat{A}'(T)V_5(T) + 3\hat{A}'(T)V_3''(T) \\
&= V_7(T) + 6\hat{A}'(T)V_5(T) + 3\hat{A}'(T)^2V_3(T)
\end{aligned}$$

where we used (3.7.16). Thus

$$\begin{aligned}
\sum_{i=0}^{7} V_i^{(7-i)}(T) &= 5V_7(T) + 10\hat{A}'(T)V_5(T) + 3\hat{A}'(T)^2V_3(T) \\
&= -384\,\hat{A}'(T)^3V_0'(T). \qquad (3.7.32)
\end{aligned}$$

Then (3.4.77) follows from (3.7.30)-(3.7.32). The proof of Proposition 3.4.13 is achieved.

Chapter 4

Multicolor image segmentation using Ambrosio-Tortorelli approximation

4.1 The segmentation problem

This section is concerned with image segmentation, which plays a very important role in many applications. The aim is to find a partition of an image into its constituent parts. As we will see, the main difficulty is that one needs to manipulate objects of different kinds: functions, domains in $\mathbb{R}^2$, and curves. We could say that segmenting an image means dividing it into its constituent parts.

Let us present the model introduced by Mumford and Shah in 1989 [70]. In this section Ω is a bounded open set of $\mathbb{R}^N$, $N = 2, 3$, and $u_0(x)$ is the initial image. Without loss of generality we can always assume that $0 \leq u_0(x) \leq 1$ a.e. $x \in \Omega$. We search for a pair (u, K), where $K \subset \Omega$ is the set of discontinuities, minimizing

$$F(u,K) = \int_{\Omega\setminus K} (u-u_0)^2 dx + \alpha \int_{\Omega\setminus K} |\nabla u|^2 dx + \beta \mathcal{H}^{N-1}(K),$$

where α and β are nonnegative constants and $\mathcal{H}^{N-1}(K)$ corresponds to $N-1$ dimensional Hausdorff measure of K, which is the most natural way of extending the notion of length to non-smooth sets and u belongs to the Sobolev space $W^{1,2}(\Omega \setminus K)$.

The difficulty in studying F is that it involves two unknowns u and K of different natures: u is a function defined on an N-dimensional space, while K is an $(N-1)$-dimensional set.

In order to apply the direct method of the calculus of variations, it is necessary to find a topology that ensures at the same time lower semicontinuity of F and compactness of the minimizing sequences. The difficulty comes from $H^{N-1}(K)$. Indeed, let E be a Borel set of $\mathbb{R}^N$ with topological boundary ∂E. It is easy to convince oneself that the map $E \to H^{N-1}(E)$. is not lower semicontinuous with respect to any compact topology.

This shows the necessity of finding another formulation of $F(u, K)$. The new formulation involves the space $BV(\Omega)$ of functions of bounded variation in Ω. The idea is to identify the set of edges K with the jump set S_u of u, which allows us to eliminate the unknown K. So the idea is to consider the functional

$$G(u) = \int_\Omega (u-u_0)^2 dx + \alpha \int_\Omega |\nabla u|^2 dx + \beta \mathcal{H}^{N-1}(S_u).$$

Now, it is tempting to minimize G on the space $BV(\Omega)$. Unfortunately, the space $BV(\Omega)$ may contain pathological non-constant functions that are continuous and have approximate differential equal to zero almost everywhere (a well-known example is the Cantor?Vitali function). For such a function v we have

$$G(v) = \int_\Omega (v-u_0)^2 dx \geq \inf_{u\in BV(\Omega)} G(u),$$

and since these pathological functions are dense in $L^2(\Omega)$, we get

$$\inf_{u\in BV(\Omega)} G(u) = 0,$$

which implies that the infimum of G cannot be achieved in $BV(\Omega)$ in general.

To avoid this phenomenon we must eliminate these pathological functions, which have the peculiarity that their distributional derivatives are measures concentrated on Cantor sets. Let us recall that the distributional derivative Du of a $BV(\Omega)$ function can be split into three mutually singular measures:

$$Du = \nabla u dx + (u^+ - u^-) n_u \mathcal{H}^{N-1}_{|S_u} + C_u,$$

where $J(u) = (u^+ - u^-) n_u \mathcal{H}^{N-1}_{|S_u}$ is the jump part and C_u the Cantor part. Following Di Giorgi [127, 126] we define $SBV(\Omega)$ as the space of special functions of bounded variation, which is the space of $BV(\Omega)$ functions such that $C_u = 0$. The natural question is now to establish the relation between the following two problems:

$$\inf_{u,K} \left\{ \begin{array}{ll} F(u,K), & u \in W^{1,2}(\Omega \setminus K) \cap L^\infty(\Omega) \\ & K \subset \Omega, \quad K \text{ closed}, \quad \mathcal{H}^{N-1} < \infty \end{array} \right\}, \tag{4.1.1}$$

$$\inf_u \left\{ \; G(u), \quad u \in SBV(\Omega) \cap L^\infty(\Omega) \; \right\}. \tag{4.1.2}$$

The answer can be found in Ambrosio [11] and is the consequence of the following theorem:

Theorem 4.1.1 (11)**.** *Let $K \subset \Omega$ be a closed set such that $\mathcal{H}^{N-1}(K) < \infty$ and let $u \in W^{1,2}(\Omega \setminus K) \cap L^\infty(\Omega)$. Then $u\hat{n}SBV(\Omega)$ and $S_u \subset K \cup L$ with $\mathcal{H}^{N-1}(L) = 0$.*

Since the Mumford-Shah functional was proposed as a solution for the segmentation problem [70, 8], many approaches for solving it have already been developed, where we can broadly categorize them into the parametric and implicit families. Due to the fact that implicit approaches enables the topological change of the solution, level-set proposals such as in [16, 76, 75, 83] have become widely accepted. Another advantage of the implicit approaches is that they can be easily extended to higher dimensions, in comparison to the parametric ones, where mesh solutions have to be usually implemented.

Our approach will start from a modified version of the one presented by Mumford and Shah (MS) and due to Ambrosio and Tortorelli [8, 5], which is a variational approach that benefits from a number of virtues. The method identifies the contours of the image and applies smoothing inside the regions defined by them. Since the computational implementation of the MS problem is far from straight-forward, mathematicians have developed different schemes to overcome the handling of the contours. We note that the model proposed by Ambrosio and Tortorelli, uses an auxiliary field that captures the presence of sharp transitions, controlling at the same time the smoothness of in the regions.

One famous proposal is based on the level-set based modification by Chan et. al [16], belonging to the implicit approach family. However, the numerical implementation of the Heaviside function and the rescaling of the level set function may increase the complexity of the whole algorithm. Another proposal corresponds to the soft Mumford-Shah segmentation, where probability-like weights are included in the similiarity term for each segmented color. However, this approach requires the solution of a non-linear system of coupled equations.

4.2 Background

4.2.1 Mumford-Shah

The Mumford-Shah functional consists of a combination of a similarity term with other regularization ones:

$$\mathcal{F}_{\mathrm{MS}}(u,\mathcal{K}) = \mathcal{F}_{\mathrm{sim}}(u) + \alpha\mathcal{F}_{\mathrm{regIm}}(u,\mathcal{K}) + \gamma\mathcal{H}^{N-1}(\mathcal{K}) \tag{4.2.1}$$

where the similarity and regularization terms are defined as:

$$\begin{aligned} \mathcal{F}_{\mathrm{sim}}(u) &= \int_{\Omega} |u-u_0|^2 \, d\Omega, & (4.2.2)\\ \mathcal{F}_{\mathrm{regIm}}(u,\mathcal{K}) &= \int_{\Omega\backslash\mathcal{K}} \|\nabla u\|^2 \, d\Omega, & (4.2.3) \end{aligned}$$

respectively. $u_0(\cdot)$ is the original image and $u(\cdot)$ the desired solution, where this last one belongs to the special bounded variation space (SBV). On the other hand, $\mathcal{H}^{N-1}(\mathcal{K})$ is the Haussdorf measure of the contour $(\mathcal{K})$.

Since the calculation of the optimal parameters u and $\mathcal{K}$ by minimizing $\mathcal{F}_{\mathrm{MS}}(u,\mathcal{K})$ is not a trivial task, approximations such as the Ambrosio-Tortorelli (AT) are usually used instead, where $\varepsilon > 0$ is a small parameter which goes to zero.

$$\mathcal{F}_{\mathrm{MSAT}}(u,v) = \mathcal{F}_{\mathrm{sim}}(u) + \alpha\mathcal{F}_{\mathrm{regAT}}(u,v) + \gamma\mathcal{F}^{\varepsilon}_{\mathrm{AT}}(v) \tag{4.2.4}$$

where

$$\mathcal{F}_{\text{regAT}}(u,v) = \int_\Omega \|\nabla u\|^2 v^2 \, d\Omega \tag{4.2.5}$$

$$\mathcal{F}^{\varepsilon}_{\text{AT}}(v) = \varepsilon \int_\Omega \|\nabla v\|^2 d\Omega + \frac{1}{4\varepsilon}\int_\Omega (v-1)^2 \, d\Omega, \tag{4.2.6}$$

in this case we note that the function $(v-1)$ plays the rol as an approximation of the indicator function of the set $\mathcal{K}$, $\chi_{\mathcal{K}}$ defined as $\chi_{\mathcal{K}}(x)=0$, if $x \notin \mathcal{K}$ and $\chi_{\mathcal{K}}(x)=1$, if $x \in \mathcal{K}$. The numerical solution of this problem is easier to implement than the original MS one, where the optimal solution of the AT converges (in the sense of Γ-convergence) to the MS as $\varepsilon \to 0$. When u is optimized with v fixed or vice-versa, the optimization problem is convex. However, when both functions are optimized at the same time, the convexivity is not guaranteed.

4.2.2 Active contours without edges

The proposed functional by Chan and Vese [16] is

$$\mathcal{F}_{\text{CV}}(c_1,c_2,\mathcal{K}) = \mathcal{F}_{\text{CVsim}}(c_1,c_2,\mathcal{K}) + \gamma\mathcal{F}_{\text{len}}(\mathcal{K}) + \nu\mathcal{F}_{\text{area}}(\mathcal{K})$$

where $\mathcal{F}_{\text{CVsim}}(c_1,c_2,\mathcal{K})$ is the similarity term, considering that inside the contour $\mathcal{K}$ the value of the function is c_1 and outside c_2. On the other hand, the $\mathcal{F}_{\text{len}}(\mathcal{K})$ measures the length of the contour $(\mathcal{K})$ and $\mathcal{F}_{\text{area}}(\mathcal{K})$ the enclosed area.

For the level-set version, the contour $(\mathcal{K})$ is replaced by a function ϕ where the contour corresponds to the zero level value: $\phi(\vec{x})=0$). Therefore, the individual terms become:

$$\begin{aligned}
\mathcal{F}_{\text{CVsim}}(c_1,c_2,\phi) &= \lambda_1 \int_\Omega (u_0-c_1)^2 H(\phi) d\Omega \\
&\quad + \lambda_2 \int_\Omega (u_0-c_2)^2 (1-H(\phi)) d\Omega \\
\mathcal{F}_{\text{len}}(\phi) &= \int_\Omega \delta(\phi)|\nabla\phi| d\Omega \\
\mathcal{F}_{\text{area}}(\phi) &= \int_\Omega H(\phi) d\Omega
\end{aligned}$$

where $H(\cdot)$ is the Heaviside or unit step function and $\delta(\cdot)$ is the Dirac delta function. These two, have to be approximated by other functions such as the sigmoid for the Heaviside, in order to achieve a better numerical convergence of the minimization implementation.

4.3 Proposed algorithm

4.3.1 Framework for the 2 color case

The functional to minimize with the Ambrosio-Tortorelli approximation will be

$$\mathcal{F}_{\text{2CAT}}(c_1, c_2, w, v) \;=\; \mathcal{F}_{\text{2Csim}}(c_1, c_2, w) + \alpha\mathcal{F}_{\text{regAT}}(w, v) + \gamma\mathcal{F}_{\text{AT}}(v) \tag{4.3.1}$$

where $F_{\text{regAT}}(w, v)$ and $F_{\text{AT}}(v)$ functionals are defined in section 4.2, and

$$\mathcal{F}_{\text{2Csim}}(c_1, c_2, w) = \sum_{i=1}^{2} \int_{\Omega} (c_i - u_0(\vec{x}))^2 p_{2,i}(\vec{x}) d\Omega, \tag{4.3.2}$$

where $p_{2,1}(\vec{x}) = w^2(\vec{x})$ and $p_{2,2}(\vec{x}) = (1 - w(\vec{x}))^2$.

For a numerical implementation of the optimization problem showed in section 4.2, the optimality conditions were discretized where a finite difference approach was used. The different parts of the functional are:

$$\begin{aligned}
\mathcal{F}_{\text{regAT}}(w) &= \frac{1}{N_d} \sum_{\vec{k}\in\Omega} \sum_{i=1}^{N_d} \frac{1}{\|\vec{e}_i\|^2} (w[\vec{k}] - w[\vec{k} + \vec{e}_i])^2 (v[\vec{k}] + v[\vec{k} + \vec{e}_i])^2 \\
\mathcal{F}^{\varepsilon}_{\text{AT}}(v) &= \frac{2\varepsilon}{N_d} \sum_{\vec{k}\in\Omega} \sum_{i=1}^{N_d} \frac{1}{\|\vec{e}_i\|^2} (v[\vec{k}] - v[\vec{k} + \vec{e}_i])^2 + \frac{1}{4\varepsilon} \sum_{\vec{k}\in\Omega} (v[\vec{k}] - 1)^2
\end{aligned} \tag{4.3.3}$$

where $\vec{k} = (k_1, k_2)$, $N_d = 4$, with $\vec{e}_1 = (1, 0)$, $\vec{e}_2 = (0, 1)$, $\vec{e}_3 = (1, 1)$ and $\vec{e}_4 = (1, -1)$.

4.3.2 Setup for the optimization scheme

In order to interate in the optimization, the derivatives of these parts with rescpect to w and v are:

$$\begin{aligned}
\frac{\partial \mathcal{F}_{\text{2CAT}}}{\partial w[\vec{k}]}(w, v) &= \frac{\partial \mathcal{F}_{\text{2Csim}}}{\partial w[\vec{k}]}(w) + \alpha \frac{\partial \mathcal{F}_{\text{regAT}}}{\partial w[\vec{k}]}(w, v) \\
\frac{\partial \mathcal{F}_{\text{MSAT}}}{\partial v[\vec{k}]}(w, v) &= \alpha \frac{\partial \mathcal{F}_{\text{regAT}}}{\partial v[\vec{k}]}(w, v) + \gamma \frac{\partial \mathcal{F}^{\varepsilon}_{\text{AT}}}{\partial v[\vec{k}]}(v),
\end{aligned}$$

where, for a faster convergence of the minimization algorithm, the conjugate gradient algorithm was implemented [1]. By defining a vector of variables $y = (w, v, \vec{c})$, and the conjugate vector as $d_y = (d_w, d_v, d_c)$, this will be updated by:

$$\begin{aligned}
(d_w[\vec{k}], d_v[\vec{k}], d_c[\vec{k}])^{t+1} &= (w^t[\vec{k}], v^t[\vec{k}], \vec{c}^{\,t}[\vec{k}]) \\
&\quad + \mu_t (d_w[\vec{k}], d_v[\vec{k}], d_c[\vec{k}])^t
\end{aligned}$$

where the coefficient μ_t is calculated by solving the minimum along the conjugate gradient direction:

$$\mu_t = \arg\min_{\mu} \{\mathcal{F}_{\text{MSAT}}(w^t + \mu d_w^t, v^t + \mu d_w^t, \vec{c}^{\,t} + \mu d_c^t)\} \tag{4.3.4}$$

corresponding to a minimization of a $4th$ order polynomial, which may be solved with a 1D Newton method. On the other hand, the updates of the conjugate gradients are

$$d^{t+1}[\vec{k}] = g^t[\vec{k}] + \beta^t d^t[\vec{k}] \tag{4.3.5}$$

where the g^t is a vector of the derivatives of the functional with respect to the different variables, such as

$$g_w^t[\vec{k}] = -\frac{\partial \mathcal{F}_{2\mathrm{CAT}}}{\partial w[\vec{k}]}(w^t, v^t),\ g_v^t[\vec{k}] = -\frac{\partial \mathcal{F}_{2\mathrm{CAT}}}{\partial v[\vec{k}]}(w^t, v^t) \tag{4.3.6}$$

and using as β scaling factors the Pollak-Ribiere version:

$$\beta^t = \frac{< g^t, g^t > - < g^t, g^{t-1} >}{< g^t, g^t >} \tag{4.3.7}$$

where $< \cdot, \cdot >$ is the sum of the dot products between each of the functions of the vector.

4.3.3 The 4 color segmentation setup

As is also pointed out in [17], the algorithm can be extended for segmenting 2^M different levels. For the particular case where M=2, we find that the whole functional becomes

$$\begin{aligned}\mathcal{F}_{4\mathrm{CAT}}(\vec{c}, \vec{w}) &= \mathcal{F}_{4\mathrm{Csim}}(\vec{c}, \vec{w}) + \alpha \mathcal{F}_{\mathrm{regAT}}(w_1, v_1) + \gamma \mathcal{F}_{\mathrm{AT}}(v_1) \\ &\quad + \alpha \mathcal{F}_{\mathrm{regAT}}(w_2, v_2) + \gamma \mathcal{F}_{\mathrm{AT}}(v_2)\end{aligned}$$

where $\mathcal{F}_{\mathrm{regAT}}(\cdot, \cdot)$ and $\mathcal{F}_{\mathrm{AT}}(\cdot)$ defined in eqs. (4.3.3) and

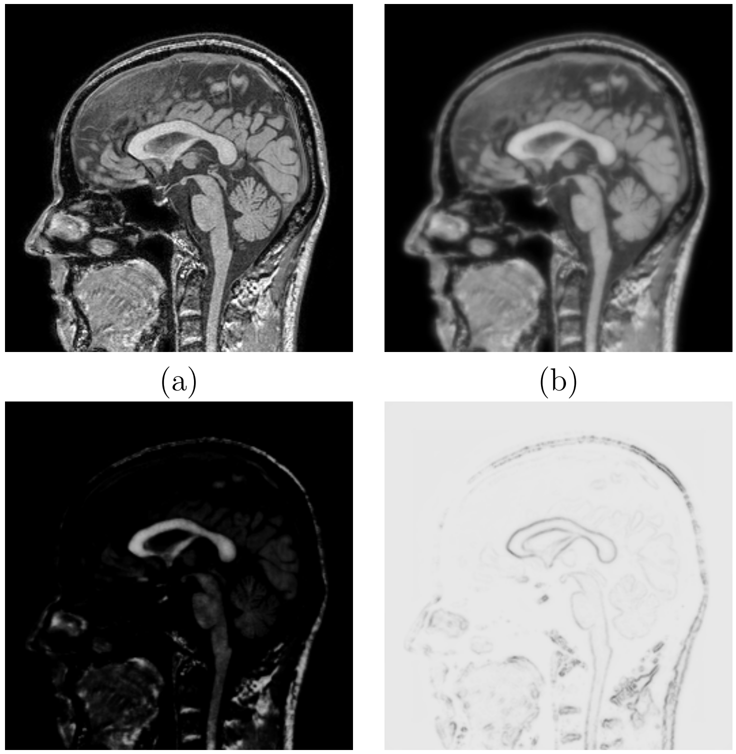
$$\mathcal{F}_{4\mathrm{Csim}}(\vec{c}, \vec{w}) = \sum_{i=1}^{4} (c_i - u_0(\vec{x}))^2 p_{4,i}(\vec{x}), \tag{4.3.8}$$

with $p_{4,1}(\vec{x}) = w_1^2 w_2^2$, $p_{4,2}(\vec{x}) = (1 - w_1)^2 w_2^2$,
$p_{4,3}(\vec{x}) = w_1^2 (1 - w_2)^2$ and $p_{4,4}(\vec{x}) = (1 - w_1)^2 (1 - w_2)^2$.

4.3.4 Further remarks

In terms of the soft Mumford-Shah segmentation [84], we can mention that for the 2 color case, we can see that the sum $p_{2,1}(\vec{x}) + p_{2,2}(\vec{x})$ does not equals to 1, except when $w(\vec{x})$ is close to 0 or 1. Nevertheless, as we will appreciate in the numerical experiments, in most of the resulting pixels, the these two extreme values are achieved. If a sum 1 of the probabilities is desired, the absolute value of the w's have to be used instead of the squares of these values.

For the 4 color case, the equivalent probabilities for the soft Mumford-Shah corresponds to the $p_{4,i}(\vec{x})$ as shown in eqs.(4.3.8), where where the sum of the unity is achieved when both, $w_1(\vec{x})$ and $w_2(\vec{x})$, are close to 0 or 1. As in the 2 color case, the sum 1 of the probabilities is achieved by using the absolute value fo the corresponding w's have to replace the squares of these.



(a) (b) (c) (d)

Figure 4.1: (a) Original Image, (b) Mumford-Shah regularized image, (c) w^2 for the 2 level segmented problem, and (d) its corresponding image contour v.

4.4 Numerical experiments

The finite difference approach was implemented in python with a finite different approach as is described in Section 4.3. Since these algorithms are nonlinear, all the input images were scaled to gray levels between 0 and 1, in order to normalize the dynamic range.

The first test involved the segmentation of a MR 2D image with the 2 color algorithm. From the results for the case where $\alpha = 5.0, \gamma = 0.2$ and $\epsilon = 0.1$, it can be appreciated that the classification was not satisfactory as is shown in figure 4.1(c), since the algorithm assumes the gray level of each class as constant. In figure 4.1(b) we can appreciate that the result using the Mumford-Shah functional with the Ambrosio-Tortorelli approach (with parameters $\alpha = 2.0, \gamma = 0.1$ and $\epsilon = 2.0$), for performing filtering, whereas the 2 color approach is suitable for segmentation.

The second numerical test was performed with another MR 2D Image, but with the 4 color algorithm. As expected, the results (with parameters $\alpha = 20.0, \gamma = 2.0$ and $\epsilon = 1.0$) were better compared to the first experiment, since these MR images have more than 4 distinct tissues. In fact, the contour images v_i present much narrower and clear transitions as shown in figures 4.2(e) and (f). With these results, it is possible to have a better tuned regularized segmentation in comparison to the 2 color case (with parameters $\alpha = 5.0, \gamma = 0.05$ and $\epsilon = 1.0$).

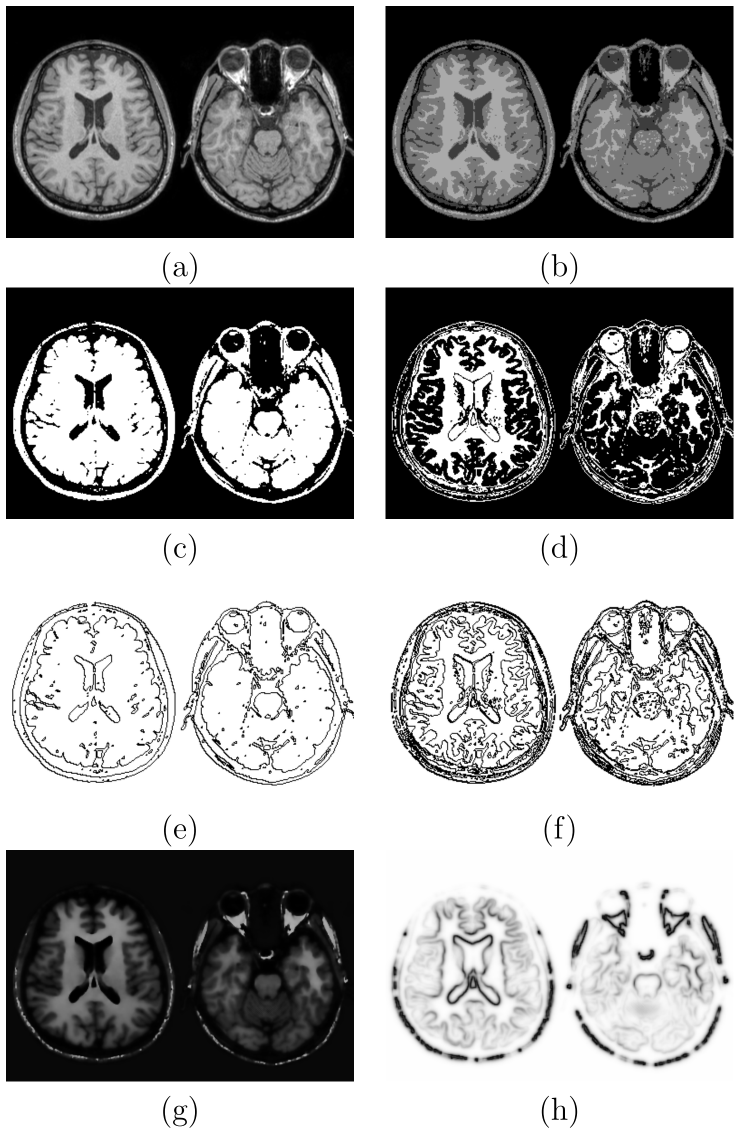


Figure 4.2: (a) Original Image, (b) 4 color (level) segmented image, (c) image region w_1, (d) w_2, (e) the contour v_1, (f) v_2; (g) 2 color segmented image w and (h) v.

Capítulo 5

Conclusiones

Los tres problemas desarrollados a lo largo de este trabajo, nos han permitido abordar diferentes aspectos del concepto de modelamiento matemático.

En el problema inverso en biología, hemos podido desarrollar el modelamiento, al incorporar variantes en el modelo y analizar el impacto de estos cambios y su interpretación.

Por otro lado, en el problema de control, hemos aplicado técnicas clásicas de control a un modelo complejo de interacción fluido-estructura, también tuvimos la oportunidad de implementar notación adecuada para incorporar fenómenos geométricos al estudio de este problema.

Finalmente, en el problema de segmentación, abordamos la implementación numérica de un algoritmo de segmentación, lo cual nos ha dado la experiencia de llevar la teoría a un nivel más aplicado, abordando otro tipo de dificultad y comprensión del modelamiento matemático.

5.1 Problema inverso en biología

En el estudio de este problema jugo un rol importante la comprensión de la interpretación biológica del modelo, para poder realizar supuestos consistentes con la biología. En esencia el estudio se centro en un problema integral de primera clase de Fredholm, es decir, la inyectividad de un operador de la forma

$$I[\rho](t) = \int_0^L \rho(x)K(t,x)dx, \quad \forall t \in [0,T], \tag{5.1.1}$$

bajo ciertos supuestos sobre el núcleos $K(t,x)$ y diferentes regularidades de ρ.

En el caso en que ρ es una función analítica, obtuvimos el Teorema 2.2.3, el cual muestra la potencia de la regularidad de ρ, al solo ser necesario que el funcional sea cero en un pequeño intervalo entorno al cero para poder obtener el resultado de identificabilidad. En cambio en el Teorema 2.2.5 no pedimos regularidad en ρ pero necesitamos que el funcional sea cero en el intervalo completo para obtener el resultado, también este resultado utiliza fuertemente la forma escogida para aproximar el núcleo, básicamente esta aproximación le brinda al núcleo la propiedad de enviar ρ a un espacio de dimensión finita, es por ellos que podemos obtener un resultado como el del Teorema 2.2.5 cuya demostración es netamente algebraica.

La reconstrucción de ρ obtenida en el Teorema 2.2.9 nos muestra el carácter difusivo del núcleo K considerado en este problema, debido a que esta reconstrucción es retrograda, es decir, el valor de ρ en un punto x depende del valor de ρ en el intervalos $(x, L]$, esto también nos muestra que este esquema propaga errores ubicados en la parte final del intervalos $[0, L]$ hacia el comienzo de este intervalo en el proceso de reconstrucción de ρ, lo cual es consistente con el mal condicionamiento descrito en los modelos numéricos desarrollados en [36, 37, 38].

La desigualdad de estabilidad obtenida en el Teorema 2.2.10, nos permite comprender la importancia de la norma utilizada y nos ofrece una herramienta para probar una desigualdad de estabilidad, para los núcleos en los cuales se pueda pasar al limite en la desigualdad (2.2.24).

En esencia este trabajo brinda herramientas para el estudio general de un problema con un núcleo difusivo, es decir, $K : [0, T] \times [0, L] \to \mathbb{R}$ el cual satisface

$$\begin{array}{ll} K(\cdot, x) & \text{es creciente } \forall x \in [0, L], \\ K(t, \cdot) & \text{es decreciente } \forall t \in [0, T]. \end{array} \tag{5.1.2}$$

Este tipo de núcleos generaliza el abordado en este trabajo. Pero varias de las ideas desarrolladas en esta tesis, como por ejemplo el uso de la transformada de Mellin en el caso que

$$K(t, x) = G\left(\frac{t}{x}\right),$$

con G una función creciente, permitirían bajo ciertos supuestos sobre la función G y la regularidad de ρ, probar la inyectividad del operador (5.1.1).

El algoritmo numérico de reconstrucción inspirado por el Teorema 2.2.10, desarrollado en la Sección 2.9 muestra un resultado consistente con los mostrados en [38], desarrollado para un núcleo general que satisfaga (5.1.2).

Por otro lado, la demostración del Teorema 2.10.1 de identificabilidad, expuesta en el caso de la aproximación polinomial del núcleo en la Sección 2.10, habré las posibilidades de obtener un resultado de estabilidad, en el caso en que ρ se escriba como suma finita de funciones

$$\psi_k(x) = \sin(\mu_k x), \tag{5.1.3}$$

con

$$\mu_k = \frac{2k+1}{2L}\pi \tag{5.1.4}$$

y utilizando una argumentación similar a la expuesta en los trabajos [29, 43, 65, 67], para construir una familia ortogonal ζ_k, tal que

$$\int_0^T \zeta_k(x) e^{\lambda_p x} dx = \delta_{kp},$$

donde λ_p es una suma a lo sumo de 8 elementos del conjunto $\{\mu_k^2 \mid \ k \in \mathbb{N}\}$.

5.2 Control de un vehículos submarino en un fluido potencial

Estudiamos la controlabilidad de un cuerpo, donde los controles que consideramos en el modelo, representan un sistema de propulsión, cada control representa una turbina, en este modelo controlamos la posición de las turbinas pero no su orientación (siempre son orientadas en la dirección perpendicular al borde del cuerpo).

Hemos logrado reducir este problema en el caso de un fluido potencial, al estudio de un sistema finito dimensional, tipo Kirchhof, el cual modela el movimiento del cuerpo en el fluido, es decir, pese a que el sistema original (3.2.1-3.2.9), que gobierna el movimiento del cuerpo, involucra una ecuación en derivadas parciales y por tanto es en esencia infinito dimensional, el estudio de la controlabilidad de este problema se ha logrado llevar al marco de un problema finito dimensional (3.3.64), con lo cual se ha podido hacer uso de las herramientas clásicas de la teoría de control, para poder estudiar el control del sistema no lineal (3.3.64).

La Proposición 3.4.1 establece la necesidad de un mínimo de 6 controles para obtener la controlabilidad del sistema linealizado (3.4.1), el cual nos entrega un resultado de controlabilidad local al rededor del origen en tiempo T del sistema (3.3.64). Esto ejemplifica la importancia del Teorema 3.4.9, el cual es obtenido mediante la aplicación del método del retorno, este método permite rescatar términos no lineales del sistema (3.3.64). Claramente la dificultad de este método radica en la construcción del recorrido entorno al origen (el cual tiene que ser solución del sistema), entorno a esta trayectoria se realiza la linealización del sistema (3.3.64). En este trabajo se trataron de seguir diferentes trayectorias pero la mayoría no lograban rescatar suficientes términos no lineales como para poder disminuir el numero de controles necesarios para obtener la controlabilidad.

Es interesante remarcar, el hecho que, la herramienta utilizada para simplificar el sistema y exhibir claramente la controlabilidad local en torno al origen en tiempo T, con solo 4 controles, es la simetría del cuerpo y de la posición de los controles sobre el cuerpo (posición de las turbinas, ver Subsección **??**), pero demasiada simetría provoca la perdida de controlabilidad, por ejemplo, si el cuerpo es una esfera (simetría total), se obtiene que el sistema 3.3.64 es no controlable, no importando el numero de controles considerados (se pierde la controlabilidad de la rotación del cuerpo), esto es concordante con la intuición del problema.

Por otro lado, siguiendo lo realizado en [44], se podría extender este resultado a un fluido invisido incluso en presencia de vorticidad, estableciendo la adecuada construcción de la solución del sistema (3.2.1-3.2.9) con la regularidad necesaria y utilizar el argumento topologico expuesto en [44] para establecer la controlabilidad incluso en presencia de vorticidad.

Otra de las lineas que se pueden continua desarrollando entorno a este trabajo, es la estimación numérica de los coeficientes de las matrices involucradas en el Teorema 3.4.9 así como el control, esto requeriría de una adecuada implementación de un método numérico para estimar ϕ la solución de la ecuación elíptica

$$\begin{aligned} \Delta\phi &= 0, \quad \text{in } \Omega, \\ \frac{\partial\phi}{\partial\hat{n}} &= g, \quad \text{on } \partial\Omega, \\ \lim_{|y|\to+\infty} \nabla\phi(y) &= 0. \end{aligned} \tag{5.2.1}$$

con Ω un domino exterior. También es interesante abordar el problema de la estabilización del sistema (3.3.64).

5.3 Segmentación de una imagen multicolor

En la búsqueda de una implementación numérica más sencilla de la solución de los problemas variacionales como el de Mumford-Shah y la segmentación multicolor, la aproximación basada en el funcional de Ambrosio-Tortorelli

$$F_\varepsilon(u,v) = \int_\Omega |u-u_0|^2 dx + \int_\Omega v^2 |\nabla u|^2 dx + \varepsilon \int_\Omega |\nabla v|^2 dx + \frac{1}{4\varepsilon}\int_\Omega (v-1)^2 dx,$$

presenta una sencilla implementación en términos de la complejidad del funcional, en comparación con la familia de los problemas llamados "active contours with level sets". Especialmente, cuando se realiza la renormalización de los conjuntos de nivel y se evita el uso de otras funciones como los polinomios.

La aproximación propuesta se puede implementar utilizando diferencias finitas y se puede extender fácilmente a más de 2 dimensiones. No es difícil extender a otro tipo de imágenes, tales como imágenes vectoriales o más de dos colores.

Bibliography

[1] L. Adams and J. L. Nazareth. *Linear and Nonlinear Conjugate Gradient-Related Methods*. SIAM, Philadelphia, 1996.

[2] S. L. Altmann. *Rotations, quaternions, and double groups*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1986.

[3] L. Ambrosio. A compactness theorem for a new class of functions of bounded variation. *Bolletino della Unione Matematica Italiana*, VII(4):857– 881, 1989.

[4] L. Ambrosio. Existence theory for a new class of variational problems. *Archive for Rational Mechanics and Analysis*, 111:291–322, 1990.

[5] L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via γ-convergence. *Communications on Pure and Applied Mathematics*, XLIII:999–1036, 1990.

[6] K. Astala and L. Päivärinta. Calderón's inverse conductivity problem in the plane. *Ann. of Math.*, 163:265–299, 2006.

[7] A. Astolfi, D. Chhabra, and R. Ortega. Asymptotic stabilization of some equilibria of an underactuated underwater vehicle. *Systems Control Lett.*, 45(3):193–206, 2002.

[8] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2006.

[9] A. M. Bloch, P. S. Krishnaprasad, J. E. Marsden, and G. Sánchez de Alvarez. Stabilization of rigid body dynamics by internal and external torques. *Automatica J. IFAC*, 28(4):745–756, 1992.

[10] I. Boekhff, E. Tareilus, J. Strotmann, and H. Breer. Rapid activation of alternative second messenger pathways in olfactory cilia from rats by different odorants. *EMBO J.*, 9:2453–2458, 1990.

[11] H. Breer, K. Raming, and J. Krieger. Signal recognition and transduction in olfactory neurons. *Biochim. Biophys. Acta.*, 1224:277–287, 1994.

[12] R. Brown and R. Torres. Uniqueness in the inverse conductivity problem for conductivities with 3/2 derivatives in L^p, $p > 2n$. *J. Fourier Analysis Appl.*, 9:1049–1056, 2003.

[13] R. Bruch and J. Teeter. Cyclic AMP links amino acid chemoreceptors to ion channels in olfactory cilia. *Chem. Senses*, 15:419–430, 1990.

[14] A. P. Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics*, pages 65–73. Soc. Brasil. Mat., Río de Janeiro, 1980.

[15] T. Chambrion and M. Sigalotti. Tracking control for an ellipsoidal submarine driven by Kirchhoff's laws. *IEEE Trans. Automat. Control*, 53(1):339–349, 2008.

[16] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10 Issue:2:266 – 277, August 2001.

[17] T. Chan and L. Vese. A level set algorithm for minimizing the mumford-shah functional in image processing. *IEEE Workshop on Variational and Level Set Methods in Computer Vision, Proceedings*, pages 161 – 168, July 2001.

[18] C. Chen, T. Nakamura, and Y. Koutalos. Cyclic AMP diffusion coefficient in frog olfactory cilia. *Biophysical J.*, 76:2861–2867, 1999.

[19] C. Conca, P. Cumsille, J. Ortega, and L. Rosier. On the detection of a moving obstacle in an ideal fluid by a boundary measurement. *Inverse Problems*, 24(4):045001, 18, 2008.

[20] C. Conca, M. Malik, and A. Munnier. Detection of a moving rigid body in a perfect fluid. *Inverse Problems*, 26:095010, 2010.

[21] J. M. Coron. Global asymptotic stabilization for controllable systems without drift. *Mathematics of Control, Signals, and Systems (MCSS)*, 5:295–312, 1992. 10.1007/BF01211563.

[22] J.-M. Coron. Contrôlabilité exacte frontière de l'équation d'Euler des fluides parfaits incompressibles bidimensionnels. *C.R. Acad. Sci. Paris*, 317:271–276, 1993.

[23] J.-M. Coron. On the controllability of 2-D incompressible perfect fluids. *J. Math. Pures Appl. (9)*, 75(2):155–188, 1996.

[24] J.-M. Coron. *Control and nonlinearity*, volume 136 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2007.

[25] J. Crank. *The Mathematics of Diffusion*. Oxford Science Publications. Clarendon Press, 1979.

[26] D. P. Dougherty, G. A. Wright, and A. C. Yew. Computational model of the camp-mediated sensory response and calcium-dependent adaptation in vertebrate olfactory receptor neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10415–10420, 2005.

[27] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, 1996.

[28] C. Fall, E. Marland, J. Wagner, and J. Tyson. *Computational Cell Biology*. New York: Springer, 2000.

[29] H. O. Fattorini and D. L. Russell. Exact controllability theorems for linear parabolic equations in one space dimension. *Arch. Ration. Mech. Anal.*, 43:272–292, 1971.

[30] E. Fernández-Cara and E. Zuazua. Control Theory: History, Mathematical Achievements and Perspectives. *Bol. Soc. Esp. Mat. Apl.*, 2003.

[31] S. Firestein, G. Shepherd, and F. Werblin. Time course of the membrane current underlying sensory transduction in salamander olfactory receptor neurones. *J. Physiol.*, Lond. 430:135–158, 1990.

[32] S. Firestein and F. Zufall. The cyclic nucleotide gated channel of olfactory receptor neurons. *Semin. Cell. Biol.*, 5(39-46), 1994.

[33] R. Flannery, D. A. French, and S. J. Kleene. Clustering of CNG channels in grass frog olfactory cilia. *Biophysical J.*, 91:179–188, 2005.

[34] T. I. Fossen. *Guidance and Control of Ocean Vehicles*. New York: Wiley, 1994.

[35] T. I. Fossen. A nonlinear unified state-space model for ship maneuvering and control in a seaway. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 15(9):2717–2746, 2005.

[36] D. A. French and D. A. Edwards. Perturbation approximation of solutions of a nonlinear inverse problem arising in olfaction experimentation. *J. Math. Biol.*, 55:745–765, 2007.

[37] D. A. French, R. J. Flannery, C. W. Groetsch, W. B. Krantz, and S. J. Kleene. Numerical approximation of solutions of a nonlinear inverse problem arising in olfaction experimentation. *Mathematical and Computer Modelling*, 43:945–956, 2006.

[38] D. A. French and C. W. Groetsch. Integral equation models for the inverse problem of biological ion channel distributions. *Journal of Physics: Conference Series*, 73(012006):1742–6596, 2007.

[39] E. D. Giorgi. Free discontinuity problems in calculus of variations. *Frontiers in pure and applied mathematics*, pages 55–62, 1991.

[40] E. D. Giorgi and L. Ambrosio. Un nuovo tipo di funzionale del calcolo delle variazioni. *Att. Accad. Naz. Lincei, Rend. Cl Sci. Fis. Mat. Nat.*, 82:199–210, 1988.

[41] E. D. Giorgi, M. Carriero, and A. Leaci. Existence theorem for a maximum problem with a free discontinuity set. *Archive for Rational Mechanics and Analysis*, 108:195–218, 1989.

[42] O. Glass. Exact boundary controllability of 3-D Euler equation. *ESAIM Control Optim. Calc. Var.*, 5:1–44 (electronic), 2000.

[43] O. Glass. A complex-analytic approach to the problem of uniform controllability of a transport equation in the vanishing viscosity limit. *Journal of Functional Analysis*, 258(3):852 – 868, 2010.

[44] O. Glass and L. Rosier. On the control of the motion of a boat. Prépublication IECN 2011/13 Prépublication IECN 2011/13, Mar. 2011.

[45] O. Glass and L. Rosier. On the control of the motion of a boat. *Math. Models Methods Appl. Sci.*, in press.

[46] O. Glass, F. Sueur, and T. Takahashi. Smoothness of the motion of a rigid body immersed in an incompressible perfect fluid. *preprint, arXiv:1003.4172*, 2011.

[47] C. Groetsch. *The theory of Tikhonov regularization for Fredholm equations of the first kind*. Research Notes in Mathematics. Pitman Advanced Pub. Program, 1984.

[48] O. Imanuvilov, G. Uhlmann, and M. Yamamoto. Global uniqueness from partial cauchy data in two dimensions. *preprint arXiv:0810.2286*, 2008.

[49] V. I. Judovič. A two-dimensional non-stationary problem on the flow of an ideal incompressible fluid through a given region. *Mat. Sb. (N.S.)*, 64 (106):562–588, 1964.

[50] R. E. Kalman, Y.-C. Ho, and K. S. Narendra. Controllability of linear dynamical systems. *Contributions to Differential Equations 1*, pages 189–213, 1963.

[51] A. V. Kazhikhov. Note on the formulation of the problem of flow through a bounded region using equations of perfect fluid. *Journal of Applied Mathematics and Mechanics*, 44:672–674, 1980.

[52] J. Keener and J. Sneyd. *Mathematical Physiology: Cellular Physiology*. Interdisciplinary Applied Mathematics. Springer, 2009.

[53] S. J. Kleene. Origin of the chloride current in olfactory transduction. *Neuron*, 11:123–132, 1993.

[54] S. J. Kleene and R. C. Gesteland. Calcium-activated chloride conductance in frog olfactory cilia. *J. Neurosci*, 11:3624–3629, 1991.

[55] S. J. Kleene and R. C. Gesteland. Transmembrane currents in frog olfactory cilia. *J. Membr. Biol.*, 120:75–81, 1991.

[56] S. J. Kleene, R. C. Gesteland, and S. H. Bryant. An electrophysiological survey of frog olfactory cilia. *J. Exp. Biol.*, 195:307–328, 1994.

[57] R. Kohn and M. Vogelius. Determining conductivity by boundary measurements. *Comm. Pure Appl. Math.*, 37:289–298, 1984.

[58] Y. Koutalos, K. Nakatani, and K.-W. Yau. Cyclic GMP diffusion coefficient in rod photoreceptor outer cyclic gmp diffusion coefficient in rod photoreceptor outer cyclic gmp diffusion coefficient in rod photoreceptor outer segments. *Biophysical J.*, 68:373–382, 1995.

[59] T. Kurahashi and K.-W. Yau. Co-existence of cationic and chloride components in odorant-induced current of vertebrate olfactory receptor cells. *Nature*, 363:71–74, 1993.

[60] H. Lamb. *Hydrodynamics*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, sixth edition, 1993. With a foreword by R. A. Caflisch [Russel E. Caflisch].

[61] N. E. Leonard. Stability of a bottom-heavy underwater vehicle. *Automatica J. IFAC*, 33(3):331–346, 1997.

[62] N. E. Leonard and J. E. Marsden. Stability and drift of underwater vehicle dynamics: mechanical systems with rigid motion symmetry. *Phys. D*, 105(1-3):130–162, 1997.

[63] B. Lindemann. Predicted profiles of ion concentrations in olfactory cilia in the steady state. *Biophys J.*, 80:1712–1721, April 2001.

[64] G. Lowe and G. Gold. Nonlinear amplification by calcium-dependent chloride channels in olfactory receptor cells. *Nature*, 366:283–286, 1993.

[65] W. A. J. Luxemburg and J. Korevaar. Entire functions and Müntz-Szász type approximation. *Transactions of the American Mathematical Society*, 157:23–37, 1971.

[66] L. Markus. Controllability of nonlinear processes. *J. Soc. Indust. Appl. Math. Ser. A Control*, 3:78–90, 1965.

[67] P. Martin, L. Rosier, and P. Rouchon. Null controllability of the structurally damped wave equation with moving point control. *ArXiv e-prints*, Nov. 2011.

[68] K. Matulef and W. N. Zagotta. Cyclic nucleotide-gated ion channels. *Annu Rev Cell Dev Biol*, 19:23–44, 2003.

[69] B. P. Menco. Qualitative and quantitative freeze-fracture studies on olfactory and nasal respiratory structures of frog, ox, rat, and dog. i. a general survey. *Cell Tissue Res.*, 207(2):183–209, 1980.

[70] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.

[71] A. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Ann. of Math.*, 143:71–96, 1996.

[72] S. P. Novikov and I. Shmel′tser. Periodic solutions of Kirchhoff equations for the free motion of a rigid body in a fluid and the extended Lyusternik-Shnirel′ man-Morse theory. I. *Funktsional. Anal. i Prilozhen.*, 15(3):54–66, 1981.

[73] J. H. Ortega, L. Rosier, and T. Takahashi. Classical solutions for the equations modelling the motion of a ball in a bidimensional incompressible perfect fluid. *M2AN Math. Model. Numer. Anal.*, 39(1):79–108, 2005.

[74] J. H. Ortega, L. Rosier, and T. Takahashi. On the motion of a rigid body immersed in a bidimensional incompressible perfect fluid. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 24(1):139–165, 2007.

[75] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2002.

[76] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[77] U. Pace, E. Hanski, Y. Salomon, and L. D. Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature*, 316:255–258, 1985.

[78] T. S. Reese. Olfactory cilia in the frog. *Journal of Cell Biology*, 25(2):209–230, May 1 1965.

[79] G. Ronnett and S. H. Snyder. Molecular messengers of olfaction. *Trends Neurosci*, 15(508-513), 1992.

[80] C. Rosier and L. Rosier. Smooth solutions for the motion of a ball in an incompressible perfect fluid. *J. Funct. Anal.*, 256(5):1618–1641, 2009.

[81] D. Schild and D. Restrepo. Transduction mechanisms in vertebrate olfactory receptor cells. *Physiol. Rev.*, 78:429–466, 1998.

[82] D. Schild and D. Restrepo. Transduction mechanisms in vertebrate olfactory receptor cells. *Physiological Reviews*, 78(2):429–466, 1998.

[83] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 2 edition, June 1999.

[84] J. Shen. A stochastic-variational model for soft mumford-shah segmentation. *Int. J. Biomedical Imaging*, 2006, 2006.

[85] S. G. Shirley, C. J. Robinson, K. Dickinson, R. Aujla, and G. H. Dodd. Olfactory adenylate cyclase of the rat. *Biochem. J.*, 240:605–607, 1986.

[86] P. Sklar, R. Anholt, and S. Snyder. The odorant-sensitive adenylate cyclase of olfactory receptor cells. *J. biol. Chem.*, 261:15538 –15543, 1986.

[87] E. D. Sontag. *Mathematical control theory*, volume 6 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1990. Deterministic finite-dimensional systems.

[88] B. L. Stevens and F. L. Lewis. *Aircraft Control and Simulation*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

[89] N. Suzuki, M. Takahata, and K. Sato. Oscillatory current responses of olfactory receptor neurons to odorants and computer simulation based on a cyclic amp transduction model. *Chemical Senses*, 27(9):789–801, 2002.

[90] J. Sylvester and G. Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Ann. of Math.*, 125:153–169, 1987.

[91] E. C. Titchmarsh. *Introduction to the Theory of Fourier Integrals*. Clarendon Press, Oxford, 1937.

[92] G. Uhlmann. Inverse boundary value problems for partial differential equations. In *Proceedings of the International Congress of Mathematicians, Vol. III (Berlin, 1998)*, volume Extra Vol. III, pages 77–86 (electronic), 1998.